جدول‌های آماری

نمایش داده‌ها را، با نظم خاص، در چند سطر و ستون یک جدول آماری می‌گویند. یک جدول آماری باید به نحوی تنظیم شود که بتوان به آسانی پاره‌ای از داشته‌های نهفته در داده‌ها را از روی آن خواند.

در کارهای آماری برای اینکه داده‌ها را خلاصه کنند، آن را در جدولی به نام جدول فراوانی تنظیم می‌کنند.

فراوانی و فراوانی نسبی

هرگاه n چیز از k نوع I_1 ، I_2 ، ... ، I_k باشد ($n \geq k \geq 2$) به ترتیب با تعدادهای f_1 ، f_2 ، ... ، f_k مشخص شده باشند، این تعدادها را فراوانی‌ها و $r_1 = \frac{f_1}{n}$ ، $r_2 = \frac{f_2}{n}$ ، ... ، $r_k = \frac{f_k}{n}$ را فراوانی‌های نسبی این چیزها می‌گوییم.

$f_i =$ تعداد xها در مقدار n

$$f_1 + f_2 + \cdots + f_k = \sum_{i=1}^{k} f_i = n \qquad 1 \leq f_i < n$$

فراوانی تجمعی و فراوانی نسبی تجمعی (یا فراوانی انباشته و فراوانی نسبی انباشته)

فراوانی تجمعی طبقه jام: $g_j = f_1 + f_2 + \cdots + f_j = \sum_{i=1}^{j} f_i$

فراوانی تجمعی نسبی طبقه jام: $S_j = r_1 + r_2 + \cdots + r_j = \sum_{i=1}^{j} r_i \qquad g_k = n \qquad S_k = 1$

توجه داشته باشید که:

الف) $g_i < g_{i+1}$ یعنی فراوانی تجمعی صعودی است.

ب) f_i فرقی ندارد که (صعودی یا نزولی) باشد. $g_{i+1} - g_i = f_{i+1}$

مثال: پزشکی ۲۰ بیمار تحت نظر دارد که گروه خونی آنها عبارتند از:

$A^+ \quad B^- \quad A^+ \quad O^- \quad O^+ \quad AB \quad O^- \quad O^+ \quad B^- \quad A^+$

$O^+ \quad O^- \quad B^- \quad O^+ \quad B^+ \quad B^+ \quad B^+ \quad A^+ \quad O^+ \quad A^+$

نخست ۶ گروه خونی A^+، B^+، B^-، AB، O^+ و O^- را به ترتیب با اعداد ۱، ۲، ۳، ۴، ۵ و ۶ متناظر می‌کنیم تا داده‌های اسمی به دست آیند. این اعداد صرفاً برای گروه‌بندی و نامگذاری است و نمی‌توان روی آن‌ها چهار عمل اصلی را انجام داد. داده‌های این مثال نمونه‌ای از داده‌های گسسته (داده‌های گروهی) می‌باشند.

	مقادیر متمایز x_i	f_i	r_i	g_i	s_i
A^+	۱	۵	۰/۲۵	۵	۰/۲۵
B^+	۲	۳	۰/۱۵	۸	۰/۴
B^-	۳	۳	۰/۱۵	۱۱	۰/۵۵
AB	۴	۱	۰/۰۵	۱۲	۰/۶
O^+	۵	۵	۰/۲۵	۱۷	۰/۸۵
O^-	۶	۳	۰/۱۵	۲۰	۱
		۲۰	۱		

مثال: فرض کنید ۲۰ عدد صحیح مثبت به طور تصادفی از بین ۱ تا ۲۰ تولید شده

۱۱ ۱۳ ۱۲ ۱۵ ۱۱ ۷ ۴ ۱۸ ۱۲ ۱۰

۱۰ ۴ ۱۵ ۴ ۷ ۱۸ ۱۱ ۱۸ ۱۲ ۲۰

x_i	f_i	r_i	g_i	s_i
۴	۳	۰/۱۵	۳	۰/۱۵
۷	۲	۰/۱	۵	۰/۲۵
۱۰	۲	۰/۱	۷	۰/۳۵
۱۱	۳	۰/۱۵	۱۰	۰/۵
۱۲	۳	۰/۱۵	۱۳	۰/۶۵
۱۳	۱	۰/۰۵	۱۴	۰/۷
۱۵	۲	۰/۱	۱۶	۰/۸
۱۸	۳	۰/۱۵	۱۹	۰/۹۵
۲۰	۱	۰/۰۵	۲۰	۱
	۲۰	۱		

الف ــ چه تعداد مشاهدات کوچکتر یا مساوی ۱۲ داریم.

با استفاده از فراوانی: $f_1 + f_2 + \dots + f_5 = 13$

با استفاده از فراوانی تجمعی: $g_3 = 13$

ب ــ چه تعداد داده در بازه (۱۱،۱۷] داریم:

پاسخ براساس فراوانی: $f_4 + f_5 + f_6 + f_7 = 3+3+1+2=9$

پاسخ براساس فراوانی تجمعی: $g_7 - g_3 = 16 - 7 = 9$

ج ــ چند درصد مشاهدات در بازه [۷،۱۴] قرار دارند.

پاسخ براساس فراوانی نسبی: $r_3 + r_4 + r_5 + r_6 = 0/1 + 0/15 + 0/15 + 0/05 = 0/45$

پاسخ براساس فراوانی نسبی تجمعی: $S_6 - S_2 = 0/7 - 0/25 = 0/45$ ۴۵ درصد مشاهدات در این بازه است.

د ــ فراوانی نسبی مشاهدات در بازه [۱۷،۱۹] را تعیین کنید.

مثال ۳ کتاب مطالعه شود.

مثال: داده های زیر اندازه های قد ۱۰۰ جوان بیست ساله در یکی از شهرهای ایران برحسب سانتیمتر که تا نزدیکترین واحد سرراست شده اند، می باشند:

۱۶۵، ۱۵۸، ۱۵۱، ۱۷۷، ۱۷۲، ۱۷۸، ۱۷۱، ۱۶۴، ۱۵۰، ۱۷۰، ۱۸۴، ۱۷۲

۱۸۲، ۱۵۹، ۱۵۴، ۱۶۲، ۱۷۰، ۱۵۳، ۱۵۹، ۱۶۹، ۱۶۴، ۱۶۶، ۱۷۲، ۱۶۹

۱۷۱، ۱۷۵، ۱۸۴، ۱۷۷، ۱۷۰، ۱۶۳، ۱۵۶، ۱۷۲، ۱۶۵، ۱۵۱، ۱۶۲، ۱۷۴

۱۶۹، ۱۷۴، ۱۷۵، ۱۷۶، ۱۷۲، ۱۶۲، ۱۶۹، ۱۷۲، ۱۶۵، ۱۵۰، ۱۵۶، ۱۶۴

۱۵۶، ۱۷۱، ۱۶۴، ۱۶۱، ۱۵۹، ۱۵۲، ۱۶۹، ۱۸۰، ۱۶۷، ۱۶۲، ۱۵۴، ۱۶۶

۱۷۷، ۱۸۴، ۱۷۵، ۱۷۱، ۱۶۴، ۱۵۰، ۱۷۰، ۱۸۴، ۱۷۲، ۱۶۵، ۱۷۰، ۱۶۳

۱۶۹، ۱۷۴، ۱۶۴، ۱۶۶، ۱۷۲، ۱۶۹، ۱۷۷، ۱۷۳، ۱۵۴، ۱۶۲، ۱۷۰، ۱۵۳

۱۶۲، ۱۵۱، ۱۵۸، ۱۵۳، ۱۶۳، ۱۵۷، ۱۶۵، ۱۶۵، ۱۷۴، ۱۷۹، ۱۷۷، ۱۷۹

۱۸۳، ۱۶۰، ۱۶۱، ۱۵۹

این داده ها نتیجه اندازه گیری با مقیاس پیوسته می باشند و آنها را داده های پیوسته می گویند.

تعداد رده $K = 1 + 3.322 \log_{10} n$ (*)

تعداد رده ها باید طوری انتخاب شود که در هر رده یک یا چند داده داشته باشیم، معمولاً تعداد رده ها نباید از ۵ کمتر یا از ۲۰ بیشتر باشد.

در مثال فوق همانطور که ذکر شد طریق مشخص کردن تعداد رده ها یا با استفاده از فرمول (*) و یا به صورت حدس و تجربه می باشد. در این مثال چون داده ها تا نزدیکترین واحد سرراست شده اند، لذا کوچکترین داده ۱۴۹٫۵ و بزرگترین داده ۱۸۴٫۵ می باشد.

$$\text{میزان تغییرپذیری داده ها} = S = \frac{\text{واحد گردشده داده ها}}{2} = \frac{1}{2} = 0.5$$

$$\min = \text{کوچکترین داده} - S = 150 - 0.5 = 149.5$$

$$\max = \text{بزرگترین داده} + S = 184 + 0.5 = 184.5$$

$\text{دامنه} = R = Max - Min = ۱۸۴٫۵ - ۱۴۹٫۵ = ۳۵$

اگر تعداد رده‌ها را ۵ در نظر بگیریم آنگاه

$W = R/K = \frac{۳۵}{۵} = ۷$ طول رده

نقطه وسط هر رده، برای نشان رده نام، که با x_i نشان می‌دهند، نماینده آن رده نامیده می‌شود.

حدود رده	x_i	f_i	r_i	g_i	s_i
۱۴۹٫۵−۱۵۶٫۵	۱۵۳	۱۰	۰/۱۰	۱۰	۰/۱۰
۱۵۶٫۵−۱۶۳٫۵	۱۶۰	۲۰	۰٫۲	۳۵	۰/۳۵
۱۶۳٫۵−۱۷۰٫۵	۱۶۷	۳۰	۰٫۳	۶۵	۰/۶۵
۱۷۰٫۵−۱۷۷٫۵	۱۷۴	۲۵	۰/۲۵	۹۰	۰٫۹
۱۷۷٫۵−۱۸۴٫۵	۱۸۱	۱۰	۰/۱	۱۰۰	۱
		۱۰۰	۱		

مثال: وزن ۴۰ تا لیسکره که به نزدیکترین عدد صحیح گرد شده اند به قرار زیر است

۵۲	۳۵	۲۴	۴۷	۳۶	۵۱	۳۴	۳۸	۴۶	۴۳
۴۷	۳۶	۳۸	۵۰	۴۷	۳۴	۴۱	۴۰	۴۲	۴۰
۲۶	۲۹	۳۰	۳۲	۳۰	۳۵	۳۷	۳۷	۴۱	۲۱
۳۱	۳۰	۲۶	۳۵	۴۵	۲۳	۴۳	۳۱	۳۴	۴۳

یک جدول فراوانی برای این داده‌ها تشکیل دهید

$K = 1 + ۳٫۳۲۲ \log_{۱۰} ۴۰ = ۶٫۳۲۲ \approx ۷$

حل:

$\delta = \frac{۱}{۲} = ۰/۵$

$min = \text{کمترین داده} - \delta = ۲۱ - ۰/۵ = ۲۰/۵$

$max = \text{بزرگترین داده} + \delta = ۵۲ + ۰/۵ = ۵۲٫۵$

$\text{دامنه} = R = max - min = ۵۲٫۵ - ۲۰/۵ = ۳۲$

$W = R/K = \frac{۳۲}{۷} = ۴٫۵۷۱۴ \approx ۵$

طول رده $w = \frac{R}{k} = \frac{۳۲}{۷} = ۴,۵۷۱۴ \simeq ۵$

رده‌ها	x_i	f_i	r_i	g_i	s_i	$f_i x_i$	$f_i x_i^2$
۲۰,۵ – ۲۵,۵	۲۳	۳	۰,۰۷۵	۳	۰,۰۷۵	۶۹	۱۵۸۷
۲۵,۵ – ۳۰,۵	۲۸	۶	۰,۱۵	۹	۰,۲۲۵	۱۶۸	۴۷۰۴
۳۰,۵ – ۳۵,۵	۳۳	۱۰	۰,۲۵	۱۹	۰,۴۷۵	۳۳۰	۱۰۸۹۰
۳۵,۵ – ۴۰,۵	۳۸	۸	۰,۲	۲۷	۰,۶۷۵	۳۰۴	۱۱۵۵۲
۴۰,۵ – ۴۵,۵	۴۳	۶	۰,۱۵	۳۳	۰,۸۲۵	۲۵۸	۱۱۰۹۴
۴۵,۵ – ۵۰,۵	۴۸	۵	۰,۱۲۵	۳۸	۰,۹۵	۲۴۰	۱۱۵۲۰
۵۰,۵ – ۵۵,۵	۵۳	۲	۰,۰۵	۴۰	۱	۱۰۶	۵۶۱۸
						۱۴۷۵	۵۶۹۶۵

با استفاده از این جدول مشخص می‌شود ۲۵٪ از تاپ‌هایی که دارای وزن بین ۳۰,۵–۳۵,۵ می‌باشند و ۴۷,۵٪ از تاپ‌هایی که دارای وزن کمتر از ۳۵,۵ می‌باشند.

مثال: داده‌های زیر یک نمونه ۵۰ تایی از اندازه نیروی پارگی نخ‌های کتانی می‌باشد

۲۱,۲	۲۸,۳	۲۷,۱	۲۵	۳۲,۷	۲۹,۵	۳۰,۲	۲۳,۹	۲۳	۲۶,۴
۲۷,۳	۳۳,۷	۲۹,۴	۲۱,۹	۲۹,۳	۱۷,۳	۲۹	۳۶,۸	۲۹,۲	۲۳,۵
۲۰,۶	۲۹,۵	۲۱,۸	۳۷,۵	۳۳,۵	۲۹,۶	۲۶,۸	۲۸,۷	۳۴,۸	۱۸,۶
۲۵,۴	۳۴,۱	۲۷,۵	۲۹,۶	۲۲,۲	۲۲,۷	۳۱,۳	۳۳,۲	۳۷	۳۸,۳
۳۶,۹	۲۴,۴	۲۸,۹	۲۴,۸	۲۸,۱	۲۵,۴	۳۴,۵	۲۳,۶	۳۸,۴	۲۴

حل:

$k = ۱ + ۳,۳۲۲ \log_{۱۰} ۵۰ = ۶,۶۴۴ \simeq ۷$

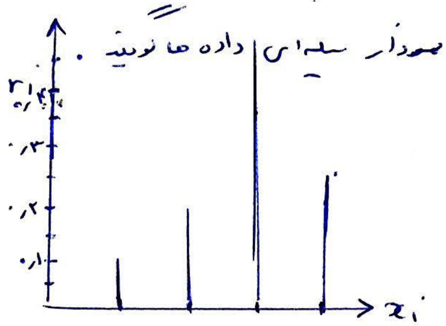
$S = \frac{۰,۱}{۲} = ۰,۰۵$ $\quad$ $\min = ۱۷,۳ - ۰,۰۵ = ۱۷,۲۵$ $\quad$ $\max = ۳۸,۴ + ۰,۰۵ = ۳۸,۴۵$

$R = ۳۸,۴۵ - ۱۷,۲۵ = ۲۱,۲$ $\quad$ $w = \frac{۲۱,۲}{۷} = ۳,۰۲۸۶ \simeq ۳,۱$

رده‌ها	x_i	f_i	r_i	g_i	s_i
۱۷,۲۵–۲۰,۳۵	۱۸,۸	۲	۰,۰۴	۲	۰,۰۴
۲۰,۳۵–۲۳,۴۵	۲۱,۹	۷	۰,۱۴	۹	۰,۱۸
۲۳,۴۵–۲۶,۵۵	۲۵	۱۰	۰,۲	۱۹	۰,۳۸
۲۶,۵۵–۲۹,۶۵	۲۸,۱	۱۷	۰,۳۴	۳۶	۰,۷۲
۲۹,۶۵–۳۲,۷۵	۳۱,۲	۳	۰,۰۶	۳۹	۰,۷۸
۳۲,۷۵–۳۵,۸۵	۳۴,۳	۶	۰,۱۲	۴۵	۰,۹
۳۵,۸۵–۳۸,۹۵	۳۷,۴	۵	۰,۱	۵۰	۱

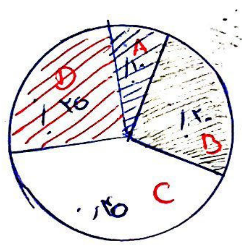
نمودارهای آماری برای داده‌های گسسته

الف - نمودار میله‌ای: در این نمودار دو محور عمود بر هم در نظر می‌گیریم و بر روی محور افقی مقادیر x_iها و بر روی محور عمودی مقادیر فراوانی نسبی x_iها را نمایش می‌دهیم. سپس از هر مقدار x_i میله‌ای به ارتفاع فراوانی نسبی x_i مربوط به آن طبقه رسم می‌کنیم. نمودار حاصل را نمودار میله‌ای داده‌ها گویند.



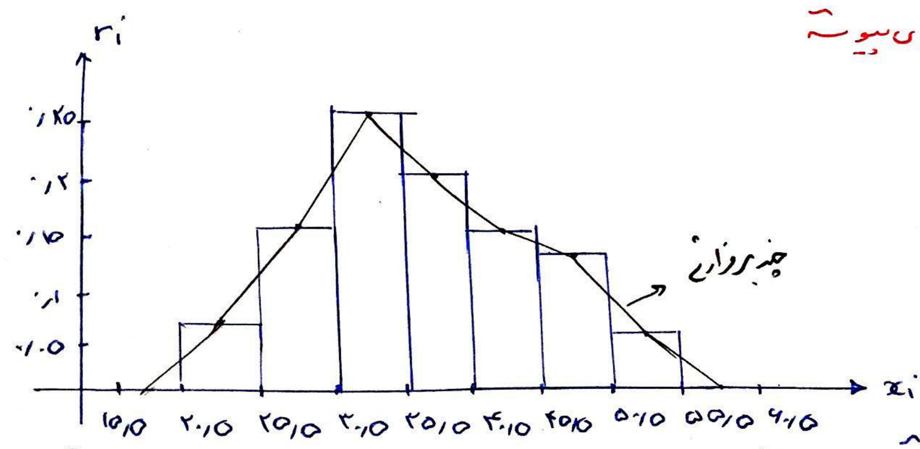
ب - نمودار دایره‌ای: در این نمودار دایره‌ای را رسم کرده و این دایره را به تعداد طبقات جدول فراوانی به قطاع‌هایی تقسیم می‌کنیم، به طوری که اندازهٔ هر قطاع متناسب با فراوانی نسبی طبقه مربوطه باشد.

برای مثال، طبقه دوم (مقدار B) دارای فراوانی نسبی ۰/۲ می‌باشد و برای این طبقه یک قطاع به اندازه $0/2 \times 360^\circ = 72^\circ$ را در دایره در نظر می‌گیریم و این عمل را برای طبقات دیگر نیز تکرار می‌کنیم.



نمودارهای آماری برای داده‌های پیوسته

الف - هیستوگرام (نمودار مستطیلی)



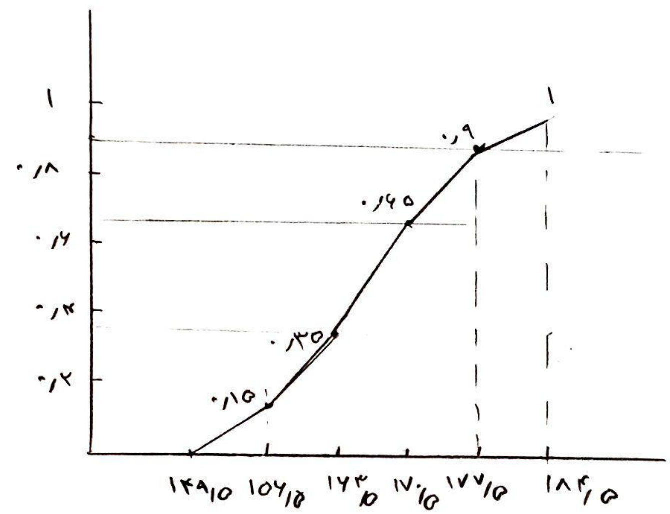
قاعده هر مستطیل روی محور افقی قرار دارد و طول آن برابر طول واقعی رده است؛ که هر چه باشد آن را یک واحد در نظر می‌گیریم و مرکز آن نماینده رده است. ارتفاع هر مستطیل برابر فراوانی نسبی مربوط به آن رده است. چون عرض هر مستطیل برابر یک واحد در نظر گرفته شده و ارتفاع هر مستطیل برابر فراوانی نسبی رده مربوطه می‌باشد پس مجموع مساحت تمام مستطیل‌های هیستوگرام

برابر یک واحد مربع است .

ب- چند بر فراوانی : اگر نقطه های وسط قاعده های بالای مستطیل های هستوگرام و نقطه های وسط رده هایی را که بلافاصله در دو انتهای هستوگرام بوده و دارای فراوانی صفر هستند به هم بپیوندیم ، یک خط شکسته به دست می آید که آن را چند بر فراوانی می نامند .

ج- چند بر فراوانی تجمعی :

اگر نقاطی را که طول آنها مرز رده ها و عرض آنها فراوانی نسبی تجمعی (انباشته) تا آن مرز باشد ، به هم بپیوندیم یک خط شکسته به دست می آید که آن را چند بر فراوانی انباشته می نامند .

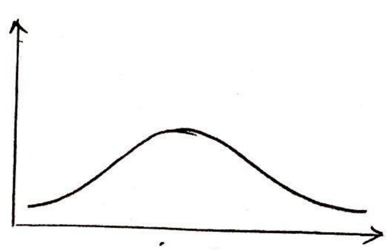


چند بر فراوانی تجمعی
(انباشته) مربوط به دمها

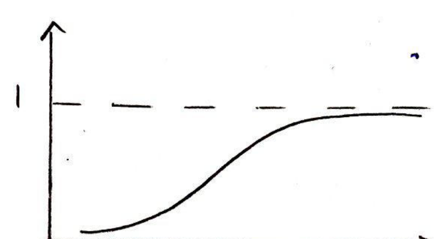
د- منحنی های فراوانی و فراوانی انباشته

اگر تعداد داده ها زیاد و طول رده ها کوچک و در نتیجه رده ها زیاد باشد ، چند بر فراوانی و چند بر فراوانی تجمعی دارای اضلاع زیاد خواهند بود و می توان برای آنها منحنی هایی منطبق کرد که به ترتیب منحنی فراوانی و منحنی فراوانی انباشته نامیده می شوند . چون مساحت تمام مستطیل های هستوگرام یک واحد مربع می باشد ، مساحت زیر منحنی فراوانی را هم یک واحد مربع فرض می کنند .

منحنی فراوانی تجمعی ، یک منحنی غیرنزولی است که عرض های نقاط آن بین صفر و یک می باشد .



منحنی فراوانی



منحنی فراوانی انباشته

هـ - منحنی فراوانی نرمال

اگر منحنی فراوانی دارای معادله مقابل

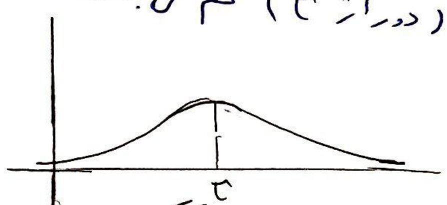
$$y = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

باشد آن را منحنی فراوانی نرمال با پارامترهای μ و σ^2 و متغیر مربوط را متغیر نرمال می‌نامند.

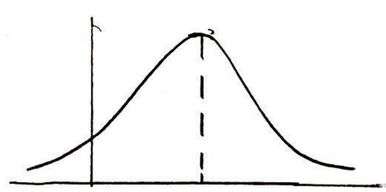
این منحنی نسبت به خط $x=\mu$ متقارن، دارای ماکسیمم در نقطه $\left(\mu , \frac{1}{\sqrt{2\pi\sigma^2}}\right)$، مجانب با محور x ها و زنگی‌شکل می‌باشد. اگر $\mu=0$ و $\sigma^2=1$، منحنی نرمال را منحنی نرمال استاندارد می‌نامند.

از عرض نقطه ماکسیمم روشن است که هرچه σ^2 کوچکتر باشد، منحنی کشیده‌تر بوده، قسمت اعظم مساحت زیر منحنی اطراف $x=\mu$ توزیع می‌شود.

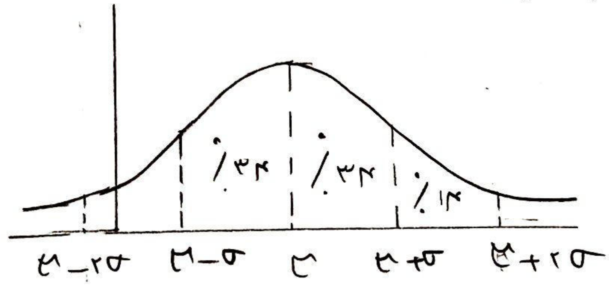
کلمه نرمال به معنی طبیعی و عادی است. معمولاً بعضی ویژگی‌های طبیعی مانند قد و وزن تقریباً دارای منحنی فراوانی نرمال هستند، یعنی تعداد افراد معمولی (اطراف μ) زیاد و تعداد افراد غیر طبیعی (دور از μ) کم می‌باشد.



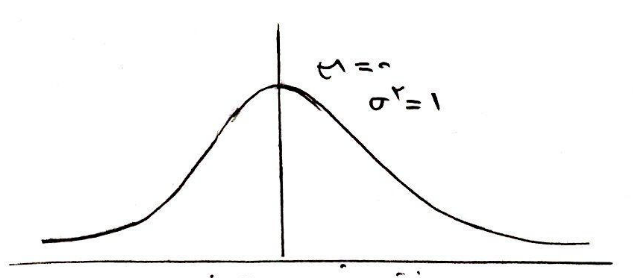
منحنی نرمال با σ^2 بزرگ



منحنی نرمال با σ^2 کوچک



نحوه توزیع مساحت زیر منحنی نرمال



منحنی نرمال استاندارد

Subject
Year: Month: Date:
Sa Su Mo Tu We Th Fr

شاخص‌های تمرکز

با استفاده از جدول فراوانی و نمودارهای آماری، می‌توان تا حدودی داشته‌های نهفته در داده‌ها را مختصر و محسوس کرد با این حال سعی می‌شود تا این داشته‌ها را به صورت یک یا چند عدد معقول درآورد، تا هم بتوان ایده‌ای کلی درباره ویژگی مورد مطالعه به دست آورد و هم نتیجه مطالعات را به سادگی گزارش داد. چنین اعدادی که معمولاً در حوالی مرکز منحنی فراوانی می‌باشند، معیارهای تمرکز نامیده می‌شوند.

۱- میانگین

میانگین، با ثبات ترین شاخص گرایش مرکزی است که با داده‌های فاصله‌ای و نسبی بکار می‌رود و مرکز ثقل داده‌ها است.

الف) میانگین حسابی

داده‌های خام:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\text{جمع همه داده‌ها}}{\text{تعداد کل داده‌ها}}$$

در جدول فراوانی:

$$\bar{x} = \frac{1}{\underbrace{n}_{\text{تعداد کل داده‌ها}}} \sum_{i=1}^{k} \underbrace{f_i}_{\text{فراوانی مطلق دسته i ام}} \underbrace{x_i}_{\text{مرکز هر دسته}}$$

Ex) در جدول فراوانی تجمعی زیر، میانگین داده‌ها کدام است.

ابتدا فراوانی مطلق هر دسته را بدست می‌آوریم.

مرکز دسته	فراوانی تجمعی	فراوانی مطلق
25	8	8
26	24	16
27	44	20
28	68	24
29	80	12

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{n} = \frac{[(25\times 8)+(26\times 16)+(27\times 20)+(28\times 24)+(29\times 12)]}{80} = 27/2$$

Ex) در جدول فراوانی زیر میانگین حسابی را بیابید.

دسته	فراوانی مطلق	مرکز دسته	$f_i x_i$
2-4	1	3	3
4-6	5	5	25
6-8	7	7	49
8-10	3	9	27
	16 (n)		104 $\rightsquigarrow \sum_{i=1}^{4} f_i x_i$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{104}{16}$$

Ex) میانگین حسابی داده‌های زیر را بدست آورید.

10, 5, 7, 3, 2, 6

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10+5+7+3+2+6}{6} = 4$$

نکته: هر تغییر در داده‌ها همان تغییر در میانگین به دنبال خواهد داشت.

Subject
Year: Month: Date:

$x_1, x_2, \dots, x_n$ ⇝ میانگین $= \bar{x}$

$\frac{x_1+a}{b}, \frac{x_2+a}{b}, \dots, \frac{x_n+a}{b}$ ⇝ میانگین $= \frac{\bar{x}+a}{b}$

$b'x_1+a, b'x_2+a, \dots, b'x_n+a$ ⇝ میانگین $= b'\bar{x}+a$

Ex) میانگین داده‌های $x_1, x_2, \dots, x_n$ برابر 3 است. میانگین داده‌های $2x_1+3, 2x_2+3, \dots, 2x_n+3$ برابر چند است

حل

با توجه به نکته بالا $b'=2$ و $a=3$ و $\bar{x}=3$ بنابراین

میانگین داده‌های جدید $= (2\times3)+3 = 9$

Ex) میانگین داده‌های یک مطالعه آماری 12 است اگر همه این داده‌ها را بر 3 تقسیم کنیم و یک واحد به آنها بیفزاییم و در مرحله آخر داده‌های جدید را دو برابر کنیم میانگین داده‌های مرحله آخر کدام است

حل) طبق مثال هر x_i به $2(\frac{x_i}{3}+1)$ تبدیل می‌شود لذا

میانگین جدید $= 2(\frac{\text{میانگین قدیم}}{3}+1) = 2(\frac{12}{3}+1) = 10$

Ex) اگر میانگین اعداد از 1 تا 100 برابر 50,5 باشد آنگاه میانگین اعداد زوج

SHAFAGH

Subject
Year: Month: Date:

از 2 تا 200 چقدر است

حل)
اگر همه اعداد در یک عدد ثابت ضرب شوند به تبع آن میانگین نیز در همان عدد ثابت ضرب می‌شود بنابراین در این مسئله میانگین دو برابر می‌شود

$50/5 \times 2 = 101$

Ex) میانگین 18 داده آماری 15/5 و میانگین 12 داده دیگر 17/25 می‌باشد. میانگین این 30 داده چقدر است

حل)

$$\bar{x} = \frac{(18 \times 15/5) + (12 \times 17/25)}{30} = \frac{486}{30} = 16/2$$

Ex) نرخ رشد میزان سرمایه‌گذاری مشتریان یک بانک در حساب‌های خود اگر در 5 سال متوالی به ترتیب برابر با 8، 25، 10، 5، 10 درصد است میانگین نرخ رشد سرمایه‌گذاری در طی این 5 سال چقدر است.

حل)

$$\bar{x} = \frac{5 + 10 + 25 + 8 + 10}{5} = \frac{58}{5} = 11/6$$

SHAFAGH

Ex) فرض کنید میانگین ۱۰ داده برابر 4 و میانگین 5 داده دیگر برابر a است. اگر میانگین کل 15 داده برابر ۱۰ باشد مقدار a را بدست آورید.

حل)

$$\frac{(4\times 10)+(5\times a)}{15}=10 \Rightarrow \frac{40+5a}{15}=10$$

$$\Rightarrow 40+5a=150 \Rightarrow 5a=110 \Rightarrow a=22$$

Subject

Year: Month: Date:

ب) میانگین وزنی

داده‌ها	وزن
x_1	w_1
x_2	w_2
$\vdots$	$\vdots$
x_n	w_n

در میانگین وزنی، وزن‌ها منعکس کننده اهمیت داده‌ها می‌باشند

$$\bar{x}_w = \sum_{i=1}^{n} w_i x_i$$ میانگین وزنی

اگر w_i ها، $w_i = \frac{f_i}{n}$

Ex) میانگین وزنی داده‌ها را بدست آورید

درس	نمره	واحد	w_i
آمار	16	3	$\frac{3}{12}$
تربیت بدنی	19	2	$\frac{2}{12}$
حسابداری	18	3	$\frac{3}{12}$
کامپیوتر	15	4	$\frac{4}{12}$
		$\sum f_i = 12$	

$$\bar{x}_w = \sum w_i x_i = (16\times\frac{3}{12}) + (19\times\frac{2}{12}) + (18\times\frac{3}{12}) + (15\times\frac{4}{12})$$

ج) میانگین هندسی

درصورتی که مقادیر x_1 و ... و x_n همگی مثبت باشند میانگین هندسی آنها با استفاده از روابط زیر محاسبه می‌شود

SHAFAGH

داده های خام :

$$\bar{X}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$$

داده های طبقه بندی شده :

$$\bar{X}_G = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}} = \sqrt[n]{x_1^{f_1} \times x_2^{f_2} \times \dots \times x_k^{f_k}}$$

البته باید توجه داشت که رابطه بالا مقدار تقریبی یا تجربی میانگین هندسی را محاسبه می کند.

از میانگین هندسی زمانی استفاده می شود که داده ها نشان دهنده تغییرات نسبی باشند. به همین دلیل از این نوع میانگین برای محاسبه متوسط رشد شاخص های نسبتی و درصدها استفاده می شود در واقع اگر تغییرات متغیر مورد نظر به ترتیب صعودی و یا نزولی بصورت نسبت یا درصد بیان شده باشد، این میانگین برای محاسبه حد متوسط آنها بکار می رود.

Ex) قیمت نوعی کالا در چهار سال گذشته برحسب تومان به ترتیب ۱۰۰، ۱۵۰، ۳۰۰ و ۸۰۰ بوده است قیمت این کالا به طور میانگین سالانه چند برابر شده است؟

حل) نسبت قیمت‌های هر سال به سال قبل عبارتند از $\frac{150}{100}$، $\frac{300}{150}$، $\frac{800}{300}$

بنابراین داریم

$$G=\sqrt[3]{\frac{150}{100}\times\frac{300}{150}\times\frac{800}{300}}=\sqrt[3]{8}=2$$

یعنی قیمت این کالا به طور متوسط سالیانه دو برابر شده است.

Ex) نسبت میزان بارندگی طی سه سال گذشته به ترتیب $\frac{20}{21}$، $\frac{15}{14}$ و $\frac{245}{432}$ بوده است میانگین این نسبت‌ها را بدست آورید

حل)

$$\bar{x}_G=\sqrt[3]{\frac{20}{21}\times\frac{15}{14}\times\frac{245}{432}}=\sqrt[3]{\frac{125}{216}}=\sqrt[3]{\frac{5^3}{6^3}}$$

$$=\frac{5}{6}$$

Ex) اگر فروش یک فروشگاه در یک سال ۸۰٪ افزایش و در سال دیگر ۸۰٪ کاهش داشته باشد، متوسط نرخ رشد در این دو سال چند درصد بوده است.

حل)
چون فروش سال اول ۸۰٪ افزایش یافته است، پس فروش آن $0/8 + 1 = 1/8$ برابر و به دلیل ۸۰٪ کاهش فروش در سال دوم، فروش آن $1 - 0/8 = 0/2$ برابر شده است. در نتیجه نرخ (ضریب) رشد در این دو سال به صورت $x_1 = 1/8$ و $x_2 = 0/2$ بوده است، پس

$$\bar{x}_G = \sqrt[2]{1/8 \times 0/2} = 0/6$$

بنابراین متوسط نرخ رشد در این دو سال $0/6 - 1 = -0/4$ درصد بوده است یعنی ۴۰٪ کاهش فروش داشته است.

توجه کنید اگر در این مثال بیان کنیم که در سال اول ۸۰٪ افزایش و در سال دوم

SHAFAGH

نیز ۷۰٪ افزایش فروش داشته است، آنگاه $\bar{x}_G = \sqrt[2]{1/8 \times 1/7} = 1/74$

در نتیجه متوسط نرخ رشد در این دو سال $0/74 = 1 - 1/74$ بدست می‌آید

یعنی ۷۴٪ افزایش فروش داشته است.

در صورتیکه محاسبه میانگین هندسی با استفاده از روابط بالا مشکل باشد

راحت‌تر است که ابتدا لگاریتم روابط بالا را محاسبه کرده سپس با معکوس

لگاریتم، مقدار میانگین هندسی را بدست آوریم. یعنی

$$\log \bar{x}_G = \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

به عبارت دیگر، لگاریتم میانگین هندسی برابر با میانگین حسابی $\log x_1$ و ... و $\log x_n$ است.

Ex) میانگین هندسی مشاهدات ۳/۶ ، ۰/۸ ، ۱۴/۱۱ را بدست آورید

حل)

$$\log \bar{x}_G = \frac{1}{3}\left(\log(3/6) + \log(0/8) + \log(14/11)\right)$$

$$= \frac{1}{3}\left(0/556 + (-0/097) + 1/15\right) = 0/536$$

$$\Rightarrow \bar{x}_G = 10^{0/536} = 3/438$$

۵) میانگین توافقی یا همساز

در صورتیکه $x_1, x_2, \ldots, x_k$ همگی غیرصفر باشند میانگین توافقی آنها را با H نشان می‌دهند.

$$\frac{1}{H}=\frac{1}{n}\sum_{i=1}^{k}\frac{f_i}{x_i} \rightsquigarrow H=\frac{n}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

به دست می‌آید. این میانگین که در عینک سنجی و مطالعه شبکه‌های برق به کار می‌رود در حقیقت برابر است با عکس میانگین حسابی $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_k}$:

Ex) یک پمپ آب یک حوضچه را در ۲۰ دقیقه و پمپ دیگر آن را در ۳۰ دقیقه پر می‌کند متوسط سرعت پمپاژ آب این دو پمپ چقدر است

حل)

$$H=\frac{1+1}{\frac{1}{20}+\frac{1}{30}}=24$$

بنابراین متوسط سرعت پمپاژ آب این دو پمپ این دو دستگاه ۲۴ دقیقه است

Ex) اتومبیلی مسافت بین دو شهر را طی کرده است. اگر $\frac{1}{3}$ مسافت بین دو شهر

SHAFAGH

را با سرعت ۱۲۰ کیلومتر در ساعت و بقیه مسیر را با سرعت ۶۰ کیلومتر در ساعت طی کرده باشد ، سرعت متوسط وی در این مسیر چقدر است ؟

حل : سرعت یک اتومبیل برابر با نسبت مسافت طی شده به زمان است . چون مسافت ثابت و با تغییر سرعت ، زمان نیز تغییر می کند ، پس از میانگین همساز استفاده می کنیم :

$$\frac{1}{H} = \frac{1}{3} \times \frac{1}{120} + \frac{2}{3} \times \frac{1}{60} \Rightarrow \frac{1}{H} = \frac{1+4}{120 \times 3} = \frac{5}{120 \times 3} = \frac{1}{3 \times 24}$$

$$\Rightarrow H = 72$$

هـ) میانگین ریشه ای مرتبه دو (میانگین درجه دوم)

$$M_2 = \left(\sum_{i=1}^{k} \frac{f_i x_i^2}{n} \right)^{\frac{1}{2}}$$

تعریف می شود ، در حقیقت برابر است با جذر میانگین حسابی

$$x_1^2, x_2^2, \ldots, x_k^2$$

مثال : میانگین درجه دوم (ریشه ای مرتبه دو) داده های ۹۶ ، ۷۲ ، ۸ ، ۲۴ ، ۶

Subject

Year: Month: Date:

را باهم مقایسه کنید.

حل)

$$\bar{x}_G = \sqrt[5]{6 \times 24 \times 8 \times 72 \times 96} = 24$$

$$M_2 = \sqrt{\frac{1}{5}(6^2+24^2+8^2+72^2+96^2)} = 54.19$$

همانطور که ملاحظه می‌شود میانگین ریشه‌ی درجه 2 از میانگین هندسی بیشتر است

رابطه بین میانگین ها

می‌توان گفت که بین چهار میانگین فوق نامساوی زیر همیشه برقرار است.

$$H \leq \bar{x}_G \leq \bar{x} \leq M_2$$

Scanned with CamScanner

Subject:
Year: Month. Date:

میانه: اگر داده ها را به صورت صعودی مرتب کنیم آنگاه عدد m را میانه داده ها گوییم که تقریباً نصف داده ها در سمت چپ و نصف داده ها در سمت راست این عدد قرار بگیرند

محاسبه میانه برای داده های گسسته

ابتدا داده های $x_1, x_2, \dots, x_n$ را به صورت صعودی $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ مرتب می کنیم. اگر n فرد باشد عدد وسط میانه است در غیر این صورت میانگین دو عدد وسط را به عنوان میانه در نظر می گیریم

$$m = \begin{cases} x_{(\frac{n+1}{2})} & \text{اگر n فرد باشد} \\ \dfrac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{اگر n زوج باشد} \end{cases}$$

Ex) میانه را برای داده ها 5، 4، 3، 7، 1، 2، 3 به دست آورید

$x_1 \; x_2 \; x_3 \; x_4 \; x_5 \; x_6 \; x_7$

ابتدا داده ها را از کوچک به بزرگ مرتب می کنیم

$\underset{x_{(1)}}{1}, \underset{x_{(2)}}{2}, \underset{x_{(3)}}{3}, \underset{x_{(4)}}{3}, \underset{x_{(5)}}{4}, \underset{x_{(6)}}{5}, \underset{x_{(7)}}{7}$ $\quad n=7$

$$m = x_{(\frac{n+1}{2})} = x_{(\frac{7+1}{2})} = x_{(\frac{8}{2})} = x_{(4)} = 3$$

PAPAZ

Ex) میانه را برای داده‌های 5, 7, 3, 3, 2, 11, 9, 8, 12, 2 بدست آورید

ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم

2, 2, 3, 3, 5, 7, 8, 9, 11, 12 $\qquad n=10 \Rightarrow \frac{n}{2}=5,\ \frac{n}{2}+1=6$

$$m=\frac{x_{(5)}+x_{(6)}}{2}=\frac{5+7}{2}=6$$

محاسبه میانه برای داده‌های پیوسته

برای محاسبه میانه در داده‌های پیوسته ابتدا باید رده‌ای که میانه در آن قرار دارد را بیابیم اولین رده‌ای که فراوانی تجمعی آن (تجمعی نسبی آن) بزرگتر یا مساوی نصف داده‌ها (0.5) است را رده شامل میانه می‌نامیم

سپس از فرمول زیر میانه را محاسبه می‌کنیم

$$m=L_i+\frac{\frac{n}{2}-g_{i-1}}{f_i}\times w$$

L_i : کران پایین رده‌ای که میانه در آن قرار دارد.

f_i : فراوانی رده میانه $\qquad g_{i-1}$: فراوانی تجمعی رده ماقبل رده میانه

w : طول رده

Subject:
Year: Month: Date:

Ex) در آخرین مثال مربوط به جدول فراوانی میانه را بدست آورید.

در این مثال $n=50$ بنابراین $\frac{n}{2}=25$ پس طبقه چهارم به عنوان طبقه میانه انتخاب می‌شود

$L_i = 26,55 \quad g_{i-1} = 19 \quad f_i = 17 \quad w = 3,1$

$$\Rightarrow m = 26,55 + \frac{\frac{50}{2} - 19}{17} \times 3,1 = 27,644$$

ویژگی‌های میانه

1- میانه به اعداد کوچک یا بزرگ حساس نیست. بنابراین بهترین شاخص است که تمرکز اعداد را در وسط توزیع نشان می‌دهد.

2- مجموع قدر مطلق انحرافات نمره‌ها از میانه کوچکتر یا مساوی مجموع قدر مطلق انحرافات نمره‌ها از هر عدد دیگری است.

نما (مد) داده‌ای که فراوانی آن بیشتر از سایر داده‌هاست را نما یا مد می‌نامیم و با حرف M نمایش می‌دهیم

محاسبه نما برای داده‌های گسسته

PAPAZ

Subject:
Year: Month: Date:

در این حالت ابتدا فراوانی داده‌ها را پیدا می‌کنیم داده‌ای که فراوانی آن بیشتر از سایر داده‌ها است را به عنوان نما در نظر می‌گیریم.

اگر دو داده‌ی غیر مجاور دارای فراوانی یکسان و بیشتر از سایر داده‌ها باشند هر دو را به عنوان نما در نظر می‌گیریم و در این حالت داده‌ها را دو نمایی می‌گوییم اگر دو داده مجاور باشند میانگین آنها را به عنوان نما در نظر می‌گیریم. اگر فراوانی داده‌ها با هم برابر باشند می‌گوییم داده‌ها بدون نما هستند.

Ex) برای داده‌های 10، 2، 3، 5، 9، 10، 14 و 9 چون فراوانی دو داده‌ی مجاور 9 و 10 از سایر داده‌ها بیشتر است پس نما برابر است با $M_o=\frac{9+10}{2}=9.5$

Ex) برای داده‌های 1، 3، 2، 1، 4، 3، 3، 5 دو داده 1 و 3 که غیر مجاور هم نیستند فراوانی مشترک آنها بیش از سایر فراوانی‌ها می‌باشد بنابراین هر دو نما هستند. این داده‌ها را دو نمایی می‌گوییم

Ex) برای داده‌های 3، 7، 11، 12، 14 و 13 چون فراوانی همه داده‌ها یکسان است پس داده‌ها بدون نما هستند

PARVAZ

محاسبه نما برای داده‌های پیوسته:

ابتدا رده‌ای که نما در آن قرار دارد را مشخص می‌کنیم (رده‌ای که فراوانی متناظر آن بیشتر از سایر رده‌هاست را به عنوان رده شامل نما در نظر می‌گیریم) سپس از فرمول زیر نما را به دست می‌آوریم.

$$M = L_m + \frac{D_1}{D_1 + D_2} \times w$$

L_m: کران پایین رده شامل نما

D_1: اختلاف فراوانی نسبی رده شامل نما و رده قبل از آن

D_2: اختلاف فراوانی نسبی رده شامل نما و رده بعد از آن

Ex) در مثال مربوط به اندازه‌های قد ۱۰۰ جوان بیست ساله نما را به دست آورید

چون بیشترین فراوانی مربوط به رده سوم است پس این رده را به عنوان رده نما انتخاب می‌کنیم

$L_m = 163/5$ $w = 7$ $D_1 = 0/10$ $D_2 = 0/05$

$M = 163/5 + \frac{0/1}{0/15} \times 7 = 168/17$

Subject:
Year: Month: Date:

Ex) در مثال مربوط به ۴۰ تا لامپ رده نما را بدست آورید

چون فراوانی رده سوم از همه بیشتر است رده نما می‌باشد

$L_m = 30.5$ $D_1 = 0.1$ $D_2 = 0.05$ $w = 5$

$$M = 30.5 + \left[\frac{0.1}{0.1 + 0.05}\right]5 = 33.833$$

چندک‌ها

عدد Q_p را که در آن $0<p<1$، چندک مرتبه p می‌نامند هرگاه تقریباً $100p$ درصد داده‌ها کمتر از آن باشند. مثلاً $Q_{0.15}$ را چندک مرتبه 0.15 می‌نامند هرگاه تقریباً 15 درصد داده‌ها کوچکتر از $Q_{0.15}$ باشند.

چندک‌ها حالت کلی‌تر از میانه می‌باشند و در حقیقت $Q_{0.5}$ همان میانه یعنی m است.

چندک‌های معروف

۱. چارک‌ها که به ازای $p = 0.25$ و 0.5 و 0.75 بدست می‌آیند آنها را به ترتیب با Q_1، Q_2، Q_3 نشان می‌دهند. Q_1 را چارک اول، Q_2 را میانه و Q_3 را چارک سوم می‌نامند. Q_1 همان میانه برای داده‌هایی که کوچکتر یا مساوی

PARVAZ

Subject:
Year: Month: Date:

m ها هستند همچنین Q_3 همان میانه‌ی داده‌های بزرگتر یا مساوی m هستند می‌باشد.

2) دهک‌ها که به ازای $p = 0/1, 0/2, \ldots, 0/9$ به دست می‌آیند و آنها را با $D_1, D_2, \ldots, D_9$ نشان می‌دهند.

3) صدک‌ها به ازای $p = 0/01, 0/02, \ldots, 0/99$ به دست می‌آیند و آنها را با $P_1, P_2, \ldots, P_{99}$ نشان می‌دهند.

به کمک این چارک‌ها می‌توان میانگین‌های گوناگونی تعریف کرد مثلاً میانگین وزنی زیر را میانگین سه‌چارکی می‌گویند

$$t = \frac{Q_1 + 2Q_2 + Q_3}{4}$$

محاسبه چندک برای داده‌های گسسته

ابتدا داده‌های $x_1, x_2, \ldots, x_n$ را به صورت غیر نزولی $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ مرتب می‌کنیم. سپس عدد $(n+1)p$ را حساب می‌کنیم اگر برابر تعداد صحیحی مانند r شد که $x_{(r)}$ را به عنوان Q_p در نظر می‌گیریم در غیر این صورت $r = [(n+1)p]$ و یک میانگین وزنی از $x_{(r)}$ و $x_{(r+1)}$ را به عنوان Q_p

PA AZ

Subject:
Year: Month: Date:

درنظر می‌گیریم یعنی

$$Q_p = (1-w)\,x_{(r)} + w\,x_{(r+1)}$$

که در آن $w = (n+1)p - r$

Ex) تعداد اتومبیل‌های تولید شده توسط یک شرکت در 18 روز کاری به شرح زیر است. چارک اول داده‌ها را حساب کنید

12, 11, 16, 15, 12, 13, 14, 12, 11, 17, 11, 13, 12, 12, 15, 12, 10, 14

حل)

10, 11, 11, 11, 12, 12, 12, 12, 12, 12, 13, 13, 14, 14, 15, 15, 16, 17

چارک اول $Q_{0/25}$

$$(n+1)p = (18+1)\times 0/25 = \overset{r}{4}/\overset{w}{75}$$

پس میانگین وزنی $x_{(4)}$ و $x_{(5)}$ چارک اول خواهد بود

$$Q_{0/25} = 0/25 \times x_{(4)} + 0/75 \times x_{(5)} = (0/25 \times 11) + (0/75 \times 12) = 11/75$$

یعنی 25 درصد روزها زیر 12 ماشین تولید می‌شود

محاسبه چندک ها برای داده های پیوسته

ابتدا باید رده ای را پیدا کنیم که فراوانی تجمعی نسبی آن بزرگتر یا مساوی p

یا فراوانی تجمعی آن بزرگتر یا مساوی np باشد

این رده، رده ای است که Q_p در آن قرار دارد سپس به کمک فرمول زیر

چندک p ام را می یابیم

فراوانی تجمعی رده قبل از رده Q_p

تعداد داده ها

$$Q_p = L_p + \frac{np - g_p}{f_p} w$$

طول رده

کران پایین رده Q_p

فراوانی رده Q_p

مثال) برای مثال مربوط به وزن قالب های کره

الف) چندک سوم را بیابید.

ب) ۹۰ درصد قالب های کره وزنشان کمتر از چه عددی است؟

حل الف) فراوانی تجمعی نسبی رده پنجم بزرگتر از 0,75 است بنابراین این

رده را به عنوان رده $Q_{0,75}$ در نظر می گیریم

$$Q_{0,75} = 40,5 + \frac{30-27}{6} \times 5 = 43$$

Subject:
Year: Month: Date:

حل ب)

$$Q_{0,9} = 45,5 + \left[\frac{36-33}{5} \times 5\right] = 48,5$$

Ex) هزینه ۱۰۰ خانوار در جدول زیر آمده است ؟ را بدست آورید $m=26$

هزینه	0-10	10-20	20-30	30-40	40-50	50-60
تعداد خانوار	10	?	25	15	?	10
g_i	10	$10+f_2$	$35+f_2$			

حل)

$$m = 26 = 20 + \frac{50-(10+f_2)}{25} \times 10 \Rightarrow 6 = \frac{40-f_2}{25} \times 10 \Rightarrow 15 = 40 - f_2$$

$$\Rightarrow f_2 = 25 \qquad f_2 + f_5 = 40 \Rightarrow f_5 = 40 - 25 = 15$$

Ex) چارک سوم را برای داده‌های زیر حساب کنید

نمرات	10-25	25-40	40-55	55-70	70-85	85-100
فراوانی	6	44	20	16	3	1
g_i	6	50	70	86	89	90

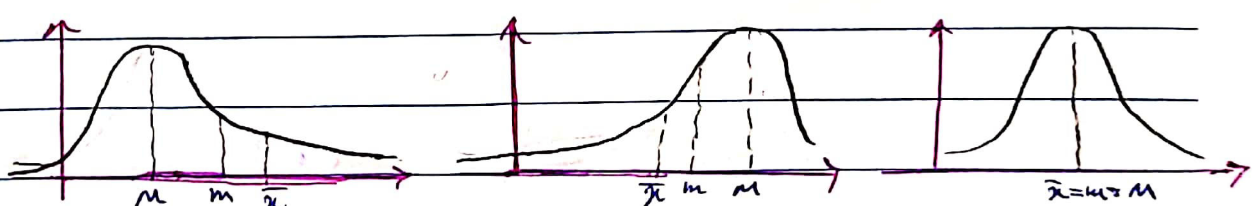
حل)

$$np = 0,75 \times 90 = 67,5 \qquad Q_p = L_p + \frac{(np - g_p)}{f_p} w$$

$$Q_p = 40 + \frac{67,5 - 50}{20} \times 15 = 40 + \frac{17,5}{20} \times 15 = 53,125$$

مقایسه میانگین، میانه و نما

میانگین برای داده‌هایی به کار می‌رود که از راه اندازه‌گیری با مقیاس فاصله‌ای یا نسبتی به دست آمده باشند. از مزایای میانگین سادگی محاسبه و تاثیر اندازه یک یک داده‌ها در این محاسبه می‌باشد. داده‌های بسیار بزرگ یا بسیار کوچک یعنی همان داده‌های پرت در میانگین اثر نامطلوب دارند و باعث می‌شوند میانگین معیار خوبی برای تمرکز نباشد و در این حالت میانه معیار بهتری می‌باشد. در ضمن اگر داده‌ها را از راه اندازه‌گیری با مقیاس‌های اسمی یا ترتیبی به دست آمده باشند باید به جای میانگین میانه و نما را به کار برد. اگر میانگین و میانه و نما بر هم منطبق باشند آنگاه داده‌ها متقارن هستند در غیر این صورت داده‌ها چوله هستند.



منحنی فراوانی چوله به راست — منحنی فراوانی چوله به چپ — متقارن

هرگاه منحنی چولگی خفیف داشته باشد بین میانگین و میانه و نما رابطه تقریبی زیر برقرار می‌باشد.

$$\bar{X} - M \simeq 3(\bar{x} - m)$$

اگر از M و m و $\bar{X}$ سه خط موازی محور yها رسم کنیم، از نظر هندسی خطی که از M رسم می‌شود از نقطه ماکزیمم منحنی فراوانی می‌گذرد خطی که m رسم می‌شود مساحت زیر منحنی فراوانی را نصف می‌کند و خطی که از $\bar{X}$ رسم می‌شود محور تعادل منحنی فراوانی را مشخص می‌سازد.

معیارهای پراکندگی

همانطور که قبلاً دیدیم داده‌ها را معمولاً به صورت یک عدد به نام معیار تمرکز خلاصه می‌کنند و قسمتی از اطلاعات موجود در آنها را در این عدد منعکس می‌سازند. ولی لازم است درباره تفاوت داده‌ها با یکدیگر و میزان پراکندگی در جمع آنها به نحوی مطالعه کرد.

مثال: فرض کنید اعداد زیر نمره‌های دو کلاس در یک آزمون ریاضی باشند

کلاس یک:

۰ ، ۶ ، ۱۲ ، ۱۰ ، ۶۲ ، ۶۳ ، ۷۰ ، ۱۰۰ ، ۱۰۰

۴۰ ، ۴۸ ، ۴۹ ، ۴۴ ، ۵۰ ، ۵۱ ، ۶۲ ، ۴۳ ، ۴۰ ، ۴۱ ، ۴۹

$$\bar{x}_1 = \bar{x}_2 = ۴۷$$

همانطور که مشاهده می‌شود در کلاس یک داده‌ها با ۴۷ تفاوت فاحشی دارند، در حالی که در کلاس دو داده‌ها اطراف ۴۷ متمرکز شده‌اند. به عبارت دیگر در کلاس یک میزان تغییرپذیری داده‌ها شدید و در کلاس دو ضعیف می‌باشد.

حال فرض کنید $x_1 , x_2 , \ldots , x_k$ یک سری داده‌های n تایی با فراوانی‌های $f_1 , f_2 , \ldots , f_k$ و میانگین $\bar{x}$ باشد. چند معیار متفاوت برای پراکندگی در زیر آورده می‌شوند.

الف: برد

$$\text{برد} \quad R = x_{(n)} - x_{(1)}$$

بزرگترین داده — کوچکترین داده

مثال: در مثال قبل برای کلاس اول $R_1 = ۱۰۰ - ۰ = ۱۰۰$ و برای کلاس دوم $R_2 = ۶۲ - ۴۰ = ۲۲$

با وجود آنکه این معیار وسعت پراکندگی را منعکس می‌کند و طرز محاسبه آن ساده می‌باشد، بیانگر خوبی برای تغییرپذیری داده‌ها نیست. زیرا این معیار فقط به بزرگترین و کوچکترین داده بستگی دارد و اگر بقیه داده‌ها تغییر کنند، تأثیری در برد ایجاد نخواهد شد. اگر تعداد داده‌ها زیاد شود، برد بزرگ شده و یا ثابت می‌ماند.

میان برد:
$$R' = \frac{x_{(1)} + x_{(n)}}{2}$$

میان برد یک نوع معیار تمرکز است. در مثال مربوط به نمرات کلاس

$$R'_1 = \frac{100 + 0}{2} = 50$$

$$R'_2 = \frac{62 + 40}{2} = \frac{102}{2} = 51$$

میانگین انحراف: معیار دیگری برای پراکندگی است

قدر مطلق $x_i - \bar{x}$ را انحراف از میانگین برای داده x_i و

$$d = \frac{\sum_{i=1}^{k} f_i |x_i - \bar{x}|}{n} \qquad (*)$$

را میانگین انحراف می‌نامند. واضح است که هرچه داده‌ها از $\bar{x}$ دورتر باشند، d بزرگتر خواهد بود.

$x_i - \bar{x}$ ممکن است مثبت یا منفی یا صفر باشد و با استفاده از تعریف $\bar{x}$ همواره داریم:

$$\sum_{i=1}^{k} f_i (x_i - \bar{x}) = 0$$

اگر میانه معیار بهتری برای تمرکز باشد، می‌توان آن را به جای میانگین در فرمول (*) به کار برد.

مثال: در مثال نمرات کلاس، میانگین انحراف‌ها برای کلاسهای یک و دو به ترتیب عبارتند از:

$$d_1 = \frac{1}{n} \sum_{i=1}^{k} f_i |x_i - \bar{x}|$$

$$\bar{x} = \frac{200 + 70 + 63 + 62 + 10 + 12 + 6}{9} = 47$$

$$d_1 = \frac{1}{9} [2 \times |100 - 47| + |70 - 47| + |63 - 47| + |62 - 47| + |10 - 47|$$
$$+ |12 - 47| + |6 - 47| + |0 - 47|]$$
$$= \frac{1}{9} \{106 + 23 + 16 + 15 + 37 + 35 + 41 + 47\} = 35/56$$

به همین ترتیب می‌توان نشان داد $d_2 = 4/91$.

چون در کلاس یک میانه $m_1 = ۴۲$ و $m_2 = ۴۸$

$۴۰, ۴۰, ۴۱, ۴۳, ۴۴, \underset{m}{۴۸}, ۴۹, ۴۹, ۵۰, ۵۱, ۶۲$

$d_1' = ۳۳/۸۸$

$$d_2' = \frac{1}{11} \sum_{i=1}^{k} f_i |x_i - m| = \frac{1}{11} \Big[۲|۴۰-۴۸| + |۴۱-۴۸| + |۴۳-۴۸|$$

$$+ |۴۴-۴۸| + |۴۸-۴۸| + ۲ \times |۴۹-۴۸| + |۵۰-۴۸| + |۵۱-۴۸| +$$

$$|۶۲-۴۸| \Big] = \frac{1}{11} \big[۱۶ + ۷ + ۵ + ۴ + ۲ + ۲ + ۳ + ۱۴ \big] = ۴/۸۲$$

میانگین انحراف ها، معیار خوبی برای نشان پراکندگی داده ها می باشد، ولی چون که به دلیل خواص ریاضی آن به علت وجود قدر مطلق، قدری پیچیده است. بنابراین به جای آن معیار دیگری به نام واریانس و یا جذر آن به نام انحراف استاندارد را به کار می برند.

واریانس و انحراف استاندارد

واریانس را که در لغت به معنی تفاوت و تغییر است از فرمول زیر به دست می آورند:

$$S^2 = \sum_{i=1}^{k} \frac{f_i (x_i - \bar{x})^2}{n}$$ میانگین مجذور انحرافات

اگر تمام داده ها به $\bar{x}$ نزدیک باشند آنگاه S^2 کوچک می شود. اگر $\forall i \; x_i = \bar{x}$ آنگاه $S^2 = 0$. بنابراین S^2 معیار خوبی برای سنجش پراکندگی و تغییرپذیری داده ها نسبت به میانگین است.

$$S^2 = \frac{1}{n} \sum_{i=1}^{k} f_i (x_i^2 + \bar{x}^2 - 2 x_i \bar{x})$$

$$= \frac{1}{n} \Big\{ \sum_{i=1}^{k} f_i x_i^2 + \bar{x}^2 \sum_{i=1}^{k} f_i - 2\bar{x} \sum_{i=1}^{k} x_i f_i \Big\}$$

$$= \frac{1}{n} \Big\{ \sum_{i=1}^{k} f_i x_i^2 + n\bar{x}^2 - 2\bar{x}(n\bar{x}) \Big\} = \frac{1}{n} \Big\{ \sum_{i=1}^{k} f_i x_i^2 - n\bar{x}^2 \Big\}$$

$$\Rightarrow S^2 = \frac{1}{n} \sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2 =$$ میانگین توان دوم داده ها − توان دوم میانگین داده ها

جذر مثبت واریانس را انحراف استاندارد می‌نامند. $s = \sqrt{s^2}$

گاهی واریانس را از فرمول

$$s_u^2 = \frac{1}{n-1} \sum_{i=1}^{k} f_i (x_i - \bar{x})^2$$

به دست می‌آورند. رابطه میان s_u^2 و s^2 به صورت زیر است

$$(n-1) s_u^2 = n s^2 \Rightarrow s_u^2 = \frac{n}{n-1} s^2$$

بدیهی است وقتی که n مثلاً $n > 100$ ، s_u^2 و s^2 تقریباً باهم برابرند.

مثال: برای داده‌های ۹، ۳، ۰، ۷، ۱ واریانس و انحراف استاندارد را حساب کنید.

$$\bar{x} = \frac{1+7+0+3+9}{5} = 4 \qquad \overline{x^2} = \frac{1}{5}[1 + 49 + 9 + 81] = 28$$

$$s^2 = 28 - (4)^2 = 28 - 16 = 12 \Rightarrow s = \sqrt{12} = 3{,}464$$

$$s_u^2 = \frac{n}{n-1} s^2 = \frac{5}{4} \times 12 = 15 \Rightarrow s_u = \sqrt{15} = 3{,}873$$

روش تبدیل یا روش کوتاه برای محاسبه میانگین و واریانس داده‌های تبدیل شده

فرض کنید $x_1, x_2, \ldots, x_k$ با فراوانی‌های $f_1, f_2, \ldots, f_k$ یک سری داده n تایی، و a و $b > 0$ دو عدد مناسب باشند. عدد a برای تغییر مبدأ اندازه‌گیری و عدد b را برای تغییر واحد اندازه‌گیری به کار می‌برند. آنگاه

$$y_i = \frac{x_i - a}{b} \quad i = 1, \ldots, k$$

را داده تبدیل شده می‌نامند. بنابراین داده‌های تبدیل شده خواهند بود $y_1, y_2, \ldots, y_k$ با فراوانی‌های $f_1, f_2, \ldots, f_k$.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{k} f_i \left(\frac{x_i - a}{b}\right) = \frac{1}{nb} \sum_{i=1}^{k} f_i x_i - \frac{a}{nb} \sum_{i=1}^{k} f_i = \frac{n\bar{x}}{nb} - \frac{a}{nb} n$$

$$= \frac{1}{b}(\bar{x} - a) \Rightarrow b\bar{y} + a = \bar{x}$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{k} f_i (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^{k} f_i \left(\frac{x_i - a}{b} - \frac{\bar{x} - a}{b} \right)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{k} f_i \left(\frac{x_i - \bar{x}}{b} \right)^2 = \frac{1}{b^2} \left[\frac{1}{n-1} \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 \right] = \frac{s_x^2}{b^2}$$

$$\Rightarrow s_x^2 = b^2 s_y^2 \quad \Rightarrow S_x = b S_y$$

مثال: در مثال اندازه پیرو پایی نخ های کتان، واریانس و انحراف استاندارد داده ها را از روش تبدیل داده ها محاسبه کنید.

x_i	f_i	$y_i = \frac{x_i - ۲۸/۱}{۳/۱}$	$f_i y_i$	$f_i y_i^2$
۱۸/۸	۲	−۳	−۶	۱۸
۲۱/۹	۷	−۲	−۱۴	۲۸
۲۵	۱۰	−۱	−۱۰	۱۰
۲۸/۱	۱۷	۰	۰	۰
۳۱/۲	۳	۱	۳	۳
۳۴/۳	۶	۲	۱۲	۲۴
۳۷/۴	۵	۳	۱۵	۴۵
	۵۰		۰	۱۲۸

$$\bar{y} = \frac{۰}{۵۰} = ۰ \qquad s_y^2 = \frac{۱}{۴۹}\left[۱۲۸ - \frac{(۰)^2}{۵۰}\right] = ۲/۶۱۲۲$$

$$\bar{x} = ۲۸/۱ + ۳/۱(۰) = ۲۸/۱ \qquad s_x^2 = (۳/۱)^2 (۲/۶۱۲۲) = ۲۵/۱۰۳$$

$$S_x = ۵/۰۱$$

داده های استاندارد

فرض کنید $x_1, x_2, \ldots, x_k$ به ترتیب با فراوانی های $f_1, f_2, \ldots, f_k$ یک سری داده n تائی با میانگین $\bar{x}$ و انحراف استاندارد S باشند. با تبدیل زیر

$$z_i = \frac{x_i - \bar{x}}{S} \qquad i = ۱, ۲, \ldots, k$$

داده های جدید $z_1, z_2, \ldots, z_k$ به ترتیب با فراوانی های $f_1, f_2, \ldots, f_k$ به نام داده های استاندارد به دست می آیند. به دیگر اگر x_i ها نمره های آزمون یک کلاس درس باشند،

z ها را نمره های استاندارد می نامند.

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{k} f_i\left(\frac{x_i-\bar{x}}{s}\right) = \frac{1}{ns}\left[\sum_{i=1}^{k} f_i x_i - \bar{x}\sum_{i=1}^{k} f_i\right]$$

$$= \frac{1}{ns}\left[n\bar{x} - \bar{x}\times n\right] = 0$$

$$S_z^2 = \frac{1}{n}\sum_{i=1}^{k} f_i\left(\frac{x_i-\bar{x}}{s}\right)^2 = \frac{1}{ns^2}\sum_{i=1}^{k} f_i(x_i-\bar{x})^2 = 1$$

از اینرو کلمه استاندارد را در مورد آنها به کار می برند. اگر متغیرها نرمال باشند، بعد از استاندارد شدن بهتر می توان آنها را باهم مقایسه و درصدهای لازم را محاسبه کرد. برای یک متغیر نرمال، $\bar{x}$ و s^2 به ترتیب تقریب های خوب برای پارامترهای منحنی نرمال یعنی μ و σ^2 می باشند.

مثال: نمره دانش آموزان یک کلاس در آزمون ریاضی، دارای میانگین ۷۲ و انحراف استاندارد ۱۵ و در آزمون فیزیک، دارای میانگین ۵۰ و انحراف استاندارد ۲۰ می باشد. اگر نمره علی در ریاضی ۶۰ و در درس فیزیک ۳۵ باشد، معلومات علی در کدام درس بیشتر است؟

چون این دو آزمون با مقیاس های مختلف به عمل آمده اند، مقایسه اعداد ۶۰ و ۳۵ معنی ندارد. اگر نمره های دو آزمون تقریباً دارای منحنی نرمال باشند، تنها بعد از استاندارد کردن می توان آنها را باهم مقایسه کرد.

$$z' = \frac{60-72}{15} = -0.8 \quad \text{نمره استاندارد ریاضی}$$

$$z'' = \frac{35-50}{20} = -0.75 \quad \text{نمره استاندارد فیزیک}$$

$\Rightarrow z'' > z'$ علی در فیزیک بهتر می تواند باشد.

ضریب تغییر: واریانس و انحراف استاندارد، به واحد اندازه گیری داده ها بستگی دارند. برای مقایسه دو سری داده بایستی از شاخص هایی استفاده کرد که به واحد اندازه گیری داده ها بستگی نداشته باشند. یکی از این شاخص ها ضریب تغییر می باشد که به صورت $v = \frac{s}{\bar{x}}$ تعریف می شود. معمولاً برحسب درصد بیان می شود.

مثال: کارخانه ای دو نوع لامپ تولید می کند. لامپ نوع اول دارای میانگین طول عمر ۲۰۰ ساعت با انحراف استاندارد ۱۱ ساعت و لامپ نوع دوم دارای میانگین طول عمر ۲۴۰ ساعت با انحراف استاندارد ۱۲ ساعت است. کدام نوع لامپ بهتر است؟

$$\bar{x}_1 = 200 \qquad S_1 = 11 \qquad V_1 = \frac{11}{200} = 0.055 \Rightarrow V_1 = \%5.5$$

$$\bar{x}_2 = 240 \qquad S_2 = 12 \qquad V_2 = \frac{12}{240} = 0.05 \Rightarrow V_2 = \%5$$

بنابراین لامپ نوع دوم بهتر است زیرا دارای میانگین طول عمر بیشتر و ضریب تغییر کمتری است.

مثال: کارخانه ای دو نوع لاستیک اتومبیل تولید می کند. برای نوع A میانگین عمر ۱۰۰۰۰ کیلومتر، با انحراف استاندارد ۲۰۰۰ کیلومتر، و برای نوع B میانگین عمر ۱۱۰۰۰ کیلومتر با انحراف استاندارد ۱۰۰۰ کیلومتر می باشد. کدام نوع لاستیک بهتر است؟

$$V_1 = \frac{2000}{10000} = \%20 \Rightarrow V_1 = \%20 \qquad \text{ضریب تغییر برای نوع A}$$

$$V_2 = \frac{1000}{11000} = 0.09 \qquad V_2 = \%9 \qquad \text{ضریب تغییر برای نوع B}$$

نوع B بهتر است، زیرا هم میانگین عمر آن بیشتر است و هم ضریب تغییر آن کوچکتر.

در این مثال اگر فرض کنیم که طول عمرها متغیرهای نرمال هستند، با محاسبه فاصله

$$(\bar{x} - 2S, \bar{x} + 2S)$$

برای هر نوع لاستیک، معلوم می شود که در حدود ۹۶٪ از لاستیک نوع A دارای طول عمری در فاصله (۶۰۰۰ , ۱۴۰۰۰) و در حدود ۹۶٪ از لاستیک نوع B، دارای طول عمری در فاصله (۹۰۰۰ , ۱۳۰۰۰) می باشند. باز هم نتیجه می شود که لاستیک نوع B بهتر است.

نیم برد چارکی

نصف برد چارکی های اول و سوم $Q = \frac{Q_3 - Q_1}{2}$

میان چارکی

$$Q' = \frac{Q_1 + Q_3}{2}$$

میان چارکی یک معیار تمرکز است. برای توزیعهای متقارن Q' درست برابر Q_2، یعنی میانه است.

$$Q = \frac{Q_3 - Q_1}{2} = \frac{Q_3 - Q_1 + Q_1}{2} - \frac{Q_1}{2} = \frac{Q_3 + Q_1}{2} - Q_1$$

$$= \frac{Q_3 - Q_3 + Q_3 - Q_1}{2} = Q_3 - \frac{Q_1 + Q_3}{2}$$

معیار Q که در حقیقت برابر فاصله Q' از Q_1 یا Q_3 است، یک نوع معیار برای سنجش پراکندگی است. به ویژه Q را در مواردی که داده های توزیع نامعلوم یا پرت، یعنی به اندازه کوچک یا به اندازه بزرگ باشند، به کار می برند.

مثال: تعداد فرزندان ۱۶ خانواده به ترتیب غیرنزولی عبارتند از:

۱، ۱، ۱، ۲، ۲، ۲، ۳، ۳، ۳، ۳، ۴، ۴، ۴، ۵، ۶، ۷

$$p = 0.25 \Rightarrow (16+1)\,0.25 = 4.25 \Rightarrow r = 4 \quad w = 0.25$$

$$Q_1 = (1-w)\,x_{(4)} + w\,x_{(5)} = 0.75 \times 2 + 2 \times 0.25 = 2$$

$$p = 0.5 \Rightarrow (16+1)\,0.5 = 17 \times 0.5 = 8.5 \Rightarrow r = 8 \quad w = 0.5$$

$$Q_2 = m = 0.5 \times x_{(8)} + 0.5\,x_{(9)} = 0.5 \times 3 + 0.5 \times 3 = 3$$

$$p = 0.75 \Rightarrow (16+1) \times 0.75 = 17 \times 0.75 = 12.75 \Rightarrow r = 12 \quad w = 0.75$$

$$Q_3 = 0.25 \times x_{(12)} + 0.75\,x_{(13)} = 0.25 \times 4 + 0.75 \times 4 = 4$$

$$\Rightarrow Q' = \frac{Q_1 + Q_3}{2} = \frac{2+4}{2} = 3$$

$$Q = \frac{Q_3 - Q_1}{2} = \frac{4-2}{2} = 1$$

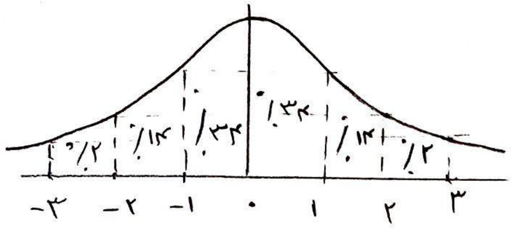
صرف نظر از داده های بالا ملاحظه می شود که Q' تقریباً مرکز توزیع و Q معیار پراکندگی می باشد.

چولگی و برجستگی

معروف‌ترین منحنی فراوانی، منحنی فراوانی نرمال استاندارد می‌باشد که معادله مختصاتی آن به صورت

$$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

است. این منحنی زنگوله‌ای شکل و از نظر متقارن، کشیدگی و پخی، تناسب و زیبایی خاصی دارد.



در عمل هیچ متغیری وجود ندارد که منحنی فراوانی آن کاملاً نرمال استاندارد باشد. اغلب منحنی فراوانی داده‌ها نامتقارن و کشیده یا پخ می‌باشند. میزان عدم نرمال بودن را با دو معیار به نام‌های چولگی و برجستگی می‌سنجند. این دو معیار به میانگین‌های مخصوصی به نام گشتاورها بستگی دارند.

گشتاور و گشتاور مرکزی داده‌ها

فرض کنید $x_1, x_2, \dots, x_k$ به ترتیب با فراوانی‌های $f_1, f_2, \dots, f_k$ یک سری داده n تایی باشند:

$$m_r = \frac{\sum_{i=1}^{k} f_i (x_i-\bar{x})^r}{n} \qquad m_r' = \frac{\sum_{i=1}^{k} f_i x_i^r}{n}$$

گشتاور r ام

گشتاور مرکزی مرتبه r ام

$$m_1' = \bar{x} \qquad m_1 = 0 \qquad m_2 = s^2$$

اگر داده‌ها نسبت به میانگین، متقارن باشد، گشتاورهای مرکزی فرد یعنی m_3، m_5، ... برابر صفر هستند.

چولگی

میزان عدم تقارن منحنی فراوانی را چولگی می‌نامند. هر کدام از فرمول‌های زیر را می‌توان به عنوان معیار چولگی به کار برد:

$b_1 = \frac{\bar{x} - M}{S}$ ضریب چولگی اول پیرسن

$b_2 = \frac{3(\bar{x} - m)}{S}$ ضریب چولگی دوم پیرسن

ضریب گشتاوری چولگی $g = \frac{m_3}{S^3}$

در فرمول‌های بالا S را بدین جهت به کار برده‌ایم، تا این ضرایب، واحد اندازه‌گیری معیّنی نداشته باشند. در صورتی که داده‌ها نسبت به میانگین متقارن باشند، ضرایب بالا برابر صفر هستند. ولی باید متوجه بود که عکس این موضوع کاملاً صحت ندارد. برحسب این که b_1، b_2 و g مثبت یا منفی باشند، منحنی فراوانی چوله به راست یا چوله به چپ خواهد بود.

چون نمامتر از میانگین و میانه، تحت تأثیر چولگی قرار می‌گیرد، از این رو ضریب b_1 را به صورت بالا تعریف کرده‌اند. ولی در عمل محاسبهٔ نما با دقت کافی مشکل می‌باشد، بنابراین b_2 را که به کمک رابطهٔ $\bar{x} - M \approx 3(\bar{x} - m)$ از روی b_1 به دست می‌آید به کار می‌برند.

مثال: طول عمر ۱۰۰ باتری اتومبیل دارای میانگین و میانه و انحراف استاندارد ۳/۵ و ۳/۴۸ و ۱/۶۵ سال می‌باشد. ضریب چولگی را محاسبه کنید.

$$b_2 = \frac{3(\bar{x} - m)}{s} = \frac{3(3/5 - 3/48)}{1/65} = 0/036$$

بنابراین منحنی فراوانی طول عمر باتری‌ها کمی چوله به راست می‌باشد.

برجستگی

میزان کشیدگی یا پخی منحنی فراوانی را نسبت به منحنی نرمال استاندارد، برجستگی آن می‌نامند.

چون برای داده‌های نرمال $\frac{m_4}{S^4}$ به عدد ۳ نزدیک می‌باشد، معیار برجستگی را از رابطهٔ

$$k = \frac{m_4}{S^4} - 3$$

به دست می‌آورند. اگر k مثبت باشد منحنی فراوانی کشیده و اگر منفی باشد پخ است. اگر k نزدیک صفر باشد، برجستگی منحنی فراوانی معمولی است.

مثال: در مثال قبل، گشتاور مرکزی چهارم برابر ۹/۲۳ است. برجستگی را محاسبه کنید.

$$K = \frac{٩,٢٤}{(١,٦٥)^{٤}} - ٣ = ١,٢٥ - ٣ = -١,٧٥$$

بنابراین منحنی فراوانی طول عمر باتری‌ها، نسبت به منحنی فراوانی نرمال استاندارد، پخ‌تر می‌باشد.

جنبه نمودار جدید

نمودارهای ساقه‌ای: (نمودار ساقه و برگ)

این نمودار معمولاً برای داده‌هایی که به صورت اعداد کوچک هستند به کار می‌رود. فرض کنید داده‌های زیر تعداد دانش‌آموزان کلاس اول ابتدایی در ۱۵ دبستان شهر A باشند. این داده‌ها به ترتیب غیر نزولی عبارتند از:

۲۲ ۲۴ ۲۵ ۳۴ ۳۵ ۴۱ ۴۱ ۴۶ ۴۶ ۴۶ ۴۷ ۴۹ ۵۴ ۵۹ ۶۰

نمودار ساقه و برگ آن به صورت زیر است:

۲	۲ ۴ ۵
۳	۴ ۵
۴	۱ ۱ ۶ ۶ ۶ ۷ ۹
۵	۴ ۹
۶	۰

الف - هر داده را دو قسمت می‌کنیم: رقم دهگان را ساقه و رقم یکان را برگ می‌نامیم. مثلاً در داده ۴۶، رقم ۴ ساقه و رقم ۶ برگ را می‌سازند.

ب - ساقه‌ها را به ترتیب صعودی از بالا به پایین طرف چپ یک خط عمودی می‌نگاریم.

ج - برگهای هر ساقه را به ترتیب غیر نزولی از چپ به راست طرف دیگر خط عمودی پهلوی آن ساقه می‌نگاریم. با مشاهده نمودار نه تنها یک داده بلکه میانه و نما (که در این مثال هر دو عدد ۴۶ هستند) دیده می‌شوند. این نمودار به هیستوگرام خوابیده شباهت دارد و نشان می‌دهد که توزیع داده‌ها متقارن نیست. اگر شماره برگهای یک ساقه زیاد باشد، آن را به دو ساقه تبدیل می‌کنند. مثلاً ساقه ۴ را به ساقه 4^{*} با برگهای ۰ تا ۴ و ساقه 4^{**} با برگهای ۵ تا ۹ تبدیل می‌کنند. اگر داده‌ها دارای ارقام اعشاری باشند، نخست آنها را به اعداد درست سرراست

می‌کنند و سپس نمودار ساقه و برگ را می‌سازند. مثلاً ۵۳٫۹۴ را به ۵۴ سرراست می‌کنند.
و آن را با ساقه ۵ و برگ ۴ نشان می‌دهند.

نمودار ساقه‌ای پشت به پشت

هنگامی که بخواهیم دو دسته داده را باهم مقایسه کنیم، دو نمودار ساقه‌ای با ساقه‌های مشترک می‌سازیم. برگ‌های یک نمودار را در سمت راست و دیگری را در سمت چپ می‌گذاریم. این نمودار را نمودار ساقه‌ای پشت به پشت می‌گویند.

مثال: داده‌های زیر تعداد دانش‌آموزان کلاس اول ابتدایی در ۱۴ دبستان شهر B است. نمودار ساقه‌ای پشت به پشت را برای مقایسه تعداد دانش‌آموزان در دو شهر A و B رسم کنید.

۹ ۸	۱	
۶ ۴ ۳	۲	۲ ۴ ۵
۴	۳	۴ ۵
۸ ۵ ۳	۴	۱ ۱ ۶ ۶ ۶ ۷ ۹
۳ ۲ ۲ ۰	۵	۴ ۹
۵	۶	۰

نمودار ساقه‌ای نقطه‌دار

همانطور که قبلاً نیز ذکر شد نمودار ساقه‌ای بر هیستوگرام برتری دارد، زیرا در نمودار ساقه‌ای یک یک داده‌ها را می‌توان دید ولی در هیستوگرام داده‌ها دیده نمی‌شوند. با این حال اگر داده‌ها با ترتیب خاصی مثلاً برحسب زمان ثبت شده باشند، نمودار ساقه‌ای نمی‌تواند این ترتیب را نشان دهد. ممکن است دو سری داده دارای یک نمودار ساقه‌ای باشند ولی زمان ثبت داده‌ها در این دو سری یکسان نباشد. آماردانی به نام هانتر در ۱۹۸۸ با طرح نمودار ساقه‌ای نقطه‌دار، این کمبود را برطرف نمود.

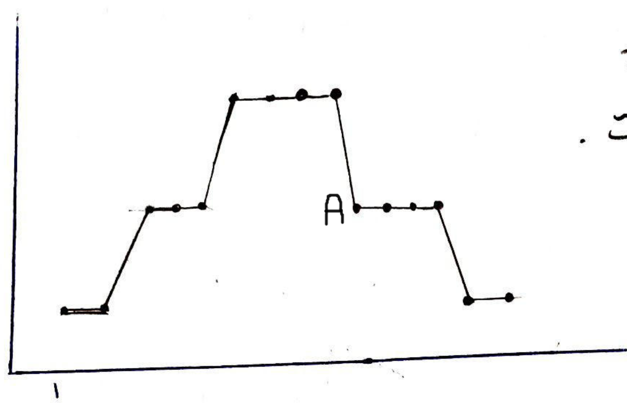
مثال: فرض کنید در شهری درجه حرارت هوای تابستان ۱۳۷۵ را هر روز اندازه گرفته‌ایم و به ترتیب

از صبح به راست اعداد زیر را برای اول تا پانزدهم مرداد ثبت کرده باشند :

۱۴ ، ۱۶ ، ۲۴ ، ۲۷ ، ۲۶ ، ۲۵ ، ۳۷ ، ۳۴ ، ۳۵ ، ۳۲ ، ۲۸ ، ۲۳ ، ۲۱ ، ۱۸ ، ۱۹

برای تعیین نمودار ساقه‌ای نقطه‌دار، نخست ساقه‌ها را میان دو خط عمودی و برگ‌ها را به ترتیب ثبت داده‌ها طرف چپ ساقه‌های مربوط می‌گذاریم. با استفاده از محور ساقه‌ها و محور زمان هر داده را با یک نقطه نشان می‌دهیم. طول این نقطه زمان ثبت داده‌ها و عرضش ساقه داده است. توجه کنید که در این نمودار برگ‌ها به ترتیب ثبت داده‌ها، نه به ترتیب صعودی قرار می‌گیرند.

۷ ۴ ۵ ۲	۳
۴ ۷ ۶ ۵ ۸ ۳ ۱	۲
۴ ۶ ۸ ۹	۱



به عنوان مثال در این شکل نقطه A می‌گوید که در دهم مرداد درجه حرارت ۲۵ است.

نمودار جعبه‌ای

این نمودار برای نشان دادن نحوه پراکندگی داده‌ها سودمند است.

روش رسم نمودار :

نخست داده‌ها را به ترتیب غیرنزولی به صورت $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$ می‌نویسیم. آنگاه آن‌ها را در پنج عدد زیر به نام خلاصه پنج عددی فشرده می‌کنیم :

۱) کوچکترین داده‌ها یعنی $x_{(1)}$

۲) بزرگترین داده‌ها یعنی $x_{(n)}$

۳) میانه داده‌ها یعنی m

۴) چارک اول داده‌ها یعنی Q_1

۵) چارک سوم داده‌ها یعنی Q_3

این ۵ عدد را با ۵ نقطه به طول‌های $x_{(n)}$ ، Q_3 ، m ، Q_1 و $x_{(1)}$ روی محور طول‌ها نشان می‌دهیم. حال مستطیلی می‌سازیم که دو ضلع عمودی آن از Q_1 و Q_3 بگذرند، و از m، خطی به موازات دو ضلع عمودی رسم می‌کنیم. از Q_1 و Q_3 دو خط دیگر به نام دنباله‌ها به ترتیب تا $x_{(1)}$ و $x_{(n)}$ امتداد می‌دهیم. بدین طریق نمودار جعبه‌ای یا نمودار جعبه‌ای دنباله‌دار به دست می‌آید. فاصله دو نقطه پایانی دنباله‌ها را با $R = x_{(n)} - x_{(1)}$ نشان می‌دهیم و آن را برد داده‌ها می‌نامیم. طول جعبه را با $Q = Q_3 - Q_1$ نشان می‌دهیم و آن را برد میان چارکی می‌نامیم. نمودار جعبه‌ای با یک نگاه نحوه پراکندگی داده‌ها را نشان می‌دهد. هر اندازه R و Q بزرگتر باشند پراکندگی بیشتر است. میزان تقارن یا عدم تقارن با این نمودار به خوبی دیده می‌شود. داده‌هائی که نسبت به سایر داده‌ها خیلی کوچکتر یا خیلی بزرگتر باشند، باعث می‌شوند که R بزرگ شود و در نتیجه دنباله چپ یا راست دراز گردد. معمولاً این نوع داده‌ها را که به داده‌های دورافتاده یا پرت شهرت دارند، کنار می‌گذارند و نمودار جعبه‌ای را با بقیه داده‌ها می‌سازند و داده‌های پرت را با نقاط جداگانه در پایان دنباله‌ها می‌گذارند. چنین نموداری را نمودار جعبه‌ای اصلاح شده می‌گویند. معمولاً اگر داده‌ای بیش از $1/5Q$ از Q_1 یا Q_3 دور باشد، آن را داده پرت محسوب می‌دارند.

مثال: داده‌های زیر نمره‌های یک آزمون ۱۰۰ تستی در یک کلاس ۲۰ نفری هستند که به ترتیب غیرنزولی تهیه شده‌اند:

۱۰ ، ۳۸ ، ۳۸ ، ۳۹ ، ۳۹ ، ۴۰ ، ۴۰ ، ۴۱ ، ۴۲ ، ۴۵
۴۷ ، ۴۸ ، ۴۹ ، ۴۹ ، ۵۰ ، ۵۱ ، ۵۱ ، ۵۳ ، ۵۳ ، ۶۱

۶۱ , ۶۳, ۶۴, ۶۵, ۶۵, ۷۰, ۷۲, ۷۳, ۷۴, ۷۵

۷۶, ۷۶, ۷۸, ۷۸, ۸۰, ۸۰, ۸۳, ۸۵, ۹۲, ۹۸

می‌خواهیم نمودار جعبه‌ای این داده‌ها را رسم کنیم. چکیده پنج عددی عبارتند از:

$x_{(1)} = ۱۰$ $\quad Q_1 = ۴۷$ $\quad m = ۶۱$ $\quad Q_3 = ۷۵$ $\quad x_{(40)} = ۹۸$

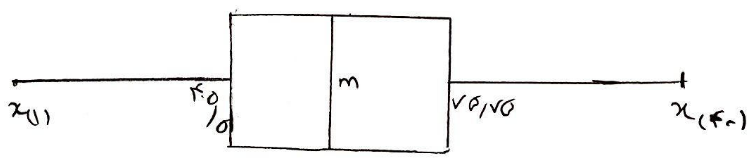
$m = \frac{۶۱ + ۶۱}{۲} = ۶۱$

$(n+1)p = ۴۱ \times \frac{1}{4} = ۱۰٫۲۵$

$Q_1 = (1 - ۰٫۲۵)x_{(10)} + ۰٫۲۵\, x_{(11)}$
$= ۰٫۷۵ \times ۴۵ + ۰٫۲۵ \times ۴۷$
$= ۴۵٫۵$

$(n+1)p = ۴۱ \times \frac{3}{4} = ۳۰٫۷۵$

$Q_3 = ۰٫۲۵\, x_{(30)} + ۰٫۷۵\, x_{(31)}$
$= ۰٫۲۵ \times ۷۵ + ۰٫۷۵ \times ۷۶$
$= ۷۵٫۷۵$



کتاب Q_1 را میانه داده‌های گرفته که از m کوچکتر یا مساوی است و Q_3 را هم "تقریباً" میانه تمام داده‌های گرفته که از m بزرگتر یا مساوی است.

چون $Q = Q_3 - Q_1 = ۲۸$ و $۱٫۵Q = ۴۲$ ؛ پس داده‌ای پرت است که از Q_1 به اندازه ۴۲ کوچکتر یا از Q_3 به اندازه ۴۲ بزرگتر باشد. در این مثال چنین داده‌ای نداریم و نیازی به اصلاح نمودار جعبه‌ای بالا نیست.

تمرین: در داده‌های زیر با استفاده از نمودار جعبه‌ای چند داده دور افتاده وجود دارد؟

۱۴, ۱۸, ۱۲, ۴۴, ۳۴, ۶۶, ۳۷, ۱۴, ۳۴, ۱۴, ۷, ۲۳, ۱۴, ۲۲, ۲۱

۱) ۰ $\quad$ ۲) ۱ $\quad$ ۳) ۲ $\quad$ ۴) ۳

مثال : داده‌های زیر عبارتند از: مدت زمانی که ۸۰ دانشجوی دانشگاه در ترم اول در امور فوق برنامه برحسب ساعت شرکت داشته‌اند

۲۱	۲۳	۲۶	۲۴	۲۴	۲۰	۱۴	۱۸	۲۴	۲۳
۲۷	۱۹	۲۰	۱۳	۱۴	۲۲	۲۰	۱۹	۱۵	۱۶
۳۱	۲۱	۱۹	۲۳	۳۲	۳۴	۲۸	۳۸	۲۲	۲۹
۱۶	۲۵	۲۱	۱۵	۲۷	۱۲	۱۸	۱۹	۲۸	۱۶
۱۱	۱۶	۲۰	۲۵	۱۸	۲۹	۲۹	۲۲	۱۷	۳۰
۱۵	۱۸	۱۷	۲۲	۲۱	۲۵	۲۴	۱۵	۱۲	۱۷
۲۲	۲۳	۲۶	۱۶	۱۵	۱۷	۱۸	۲۳	۲۰	۲۱
۱۰	۲۰	۱۵	۱۷	۱۹	۲۳	۲۰	۱۸	۱۶	۱۱

نمودار ساقه و برگ را رسم کنید:

۱	۰ ۱ ۱ ۲ ۲ ۳ ۴ ۴ ۵ ۵ ۵ ۵ ۵ ۵
۱*	۶ ۶ ۶ ۶ ۶ ۶ ۷ ۷ ۷ ۷ ۷ ۸ ۸ ۸ ۸ ۸ ۸ ۹ ۹ ۹ ۹ ۹
۲	۰ ۰ ۰ ۰ ۰ ۰ ۰ ۱ ۱ ۱ ۱ ۱ ۲ ۲ ۲ ۲ ۲ ۳ ۳ ۳ ۳ ۳ ۴ ۴ ۴ ۴ ۵ ۵ ۵
۲*	۶ ۶ ۷ ۷ ۸ ۸ ۹ ۹ ۹
۳	۰ ۱ ۲ ۴ ۸

مثال : سازمان حفاظت محیط زیست کشور آمریکا، آزمونی را بر روی اتومبیل‌های مدل جدید تولیدی کارخانه‌های گوناگون کشور انجام می‌دهد. یکی از موارد پژوهش میزان مصرف سوخت برحسب مسافت پیموده شده‌ی (برحسب مایل) هر اتومبیل در ازای مصرف یک گالن بنزین است. نتیجه‌ی این آزمایش بر روی ۱۰۰ اتومبیل مدل جدید کارخانه‌ای به شرح زیر گزارش شده است:

۳۶٫۷	۴۲٫۱	۳۷٫۲	۳۰	۳۶٫۸	۴۴٫۹	۳۷٫۱	۳۶٫۹	۴۱	۳۶٫۳
۳۳٫۶	۳۷٫۵	۳۷٫۴	۳۳٫۲	۳۶٫۵	۳۲٫۹	۳۶٫۶	۴۱٫۲	۳۷٫۳	۳۲٫۷
۳۴٫۲	۴۰	۳۷٫۷	۳۷٫۷	۳۶٫۴	۴۰٫۲	۳۳٫۹	۳۷٫۶	۳۶٫۵	۴۰٫۵

۳۵٫۱	۳۵٫۶	۳۹٫۷	۳۸٫۳	۳۸٫۲	۳۵٫۹	۳۷٫۹	۳۶	۳۷٫۹	۳۶٫۲
۳۹٫۷	۳۸٫۸	۳۴٫۴	۳۵٫۳	۳۹٫۴	۳۸٫۶	۳۴٫۸	۳۵٫۵	۳۹	۳۸٫۵
۳۹٫۴	۳۸٫۴	۳۸٫۲	۳۶٫۱	۳۶٫۶	۴۰٫۵	۳۶٫۴	۳۲٫۵	۳۶٫۸	۳۶٫۳
۳۵٫۸	۳۹	۳۸٫۷	۳۷	۳۷٫۶	۳۷	۳۳٫۱	۳۷٫۳	۳۱٫۸	۴۱
۳۴٫۵	۳۶٫۷	۳۵٫۶	۳۵٫۹	۳۷٫۸	۳۷٫۱	۳۷٫۴	۴۰٫۷	۳۷٫۲	۳۷
۳۹٫۵	۳۴٫۸	۳۵٫۲	۳۸	۴۰٫۱	۳۳٫۹	۳۷	۳۶٫۷	۴۰٫۳	۳۷٫۱
۳۶٫۹	۳۸٫۱	۳۵	۳۶٫۸	۳۴	۳۹٫۸	۳۳٫۸	۳۲٫۹	۳۶٫۹	۳۹٫۹

برای این داده‌ها فاصله‌های $(\bar{x}-s, \bar{x}+s)$ ، $(\bar{x}-2s, \bar{x}+2s)$ و $(\bar{x}-3s, \bar{x}+3s)$ را مشخص کرده آنگاه تعداد و درصد داده‌های متعلق به هر فاصله را مشخص کنید.

حل: با توجه به اینکه داده‌ها با یک رقم اعشار بیان شده‌اند پس واحد ۰٫۱ است و چون کوچکترین داده عدد ۳۰ می‌باشد، بنابراین کران پایین رده‌ی اول برابر است با ۲۹٫۹۵ = ۰٫۰۵ − ۳۰ و دامنه تغییرات برابر است با $R = ۴۴٫۹ - ۳۰ = ۱۴٫۹$

$k = ۱۰$ اختیار می‌کنیم

$w = \frac{R}{k} = \frac{۱۴٫۹}{۱۰} = ۱٫۴۹ \approx ۱٫۵$

رده	f_i	r_i	g_i	s_i	x_i	$f_i x_i$	$f_i x_i^2$
۲۹٫۹۵-۳۱٫۴۵	۱	۰٫۰۱	۱	۰٫۰۱	۳۰٫۷	۳۰٫۷	۹۴۲٫۴۹
۳۱٫۴۵-۳۲٫۹۵	۵	۰٫۰۵	۶	۰٫۰۶	۳۲٫۲	۱۶۱	۵۱۸۴٫۲
۳۲٫۹۵-۳۴٫۴۵	۹	۰٫۰۹	۱۵	۰٫۱۵	۳۳٫۷	۳۰۳٫۳	۱۰۲۲۱٫۲۱
۳۴٫۴۵-۳۵٫۹۵	۱۴	۰٫۱۴	۲۹	۰٫۲۹	۳۵٫۲	۴۹۲٫۸	۱۷۳۴۶٫۵۶
۳۵٫۹۵-۳۷٫۴۵	۳۳	۰٫۳۳	۶۲	۰٫۶۲	۳۶٫۷	۱۲۱۱٫۱	۴۴۴۴۷٫۳۷
۳۷٫۴۵-۳۸٫۹۵	۱۸	۰٫۱۸	۸۰	۰٫۸	۳۸٫۲	۶۸۷٫۶	۲۶۲۶۶٫۳۲
۳۸٫۹۵-۴۰٫۴۵	۱۲	۰٫۱۲	۹۲	۰٫۹۲	۳۹٫۷	۴۷۶٫۴	۱۸۹۱۳٫۰۸
۴۰٫۴۵-۴۱٫۹۵	۶	۰٫۰۶	۹۸	۰٫۹۸	۴۱٫۲	۲۴۷٫۲	۱۰۱۸۴٫۶۴
۴۱٫۹۵-۴۳٫۴۵	۱	۰٫۰۱	۹۹	۰٫۹۹	۴۲٫۷	۴۲٫۷	۱۸۲۳٫۲۹
۴۳٫۴۵-۴۴٫۹۵	۱	۰٫۰۱	۱۰۰	۱	۴۴٫۲	۴۴٫۲	۱۹۵۳٫۶۴
	۱۰۰	۱				۳۶۹۷	۱۳۷۲۸۲٫۸

$\bar{x} = ۳۶۹۷/۱۰۰ = ۳۶٫۹۷$

$$s_u^2 = \frac{۱۳۷۲۸۲٫۸ - \frac{(۳۶۹۷)^2}{۱۰۰}}{۱۰۰ - ۱} = ۶٫۱۱ \Rightarrow s_u = \sqrt{۶٫۱۱} = ۲٫۴۷$$

$$(\bar{x} - s , \bar{x} + s) = (۳۴٫۵ , ۳۹٫۴۴)$$

$$(\bar{x} - ۲s , \bar{x} + ۲s) = (۳۲٫۰۳ , ۴۱٫۹۱)$$

$$(\bar{x} - ۳s , \bar{x} + ۳s) = (۲۹٫۵۶ , ۴۴٫۳۸)$$

ملاحظه می‌شود که ۶۷ داده به فاصله $(\bar{x} - s , \bar{x} + s)$ تعلق دارد. همچنین ۹۶ داده به فاصله‌ی $(\bar{x} - ۲s , \bar{x} + ۲s)$ و سرانجام ۹۹ داده به فاصله $(\bar{x} - ۳s , \bar{x} + ۳s)$ تعلق دارد.

دلیل این امر آن است که توزیع فراوانی داده‌ها ناقوسی شکل است.

تمرین: یک نمونه ۲۵ تایی داده به شرح زیر گزارش شده است

۷ ۹ ۷ ۷ ۱۲ ۶ ۱۱ ۷ ۷ ۱۰

۶ ۹ ۵ ۷ ۱۰ ۱۱ ۱۰ ۹ ۷ ۸

۸ ۸ ۱۰ ۹ ۶

الف- $\bar{x}$، s^2 و s را محاسبه کنید.

ب- مقدار داده‌های متعلق به فاصله‌های $(\bar{x} - s , \bar{x} + s)$، $(\bar{x} - ۲s , \bar{x} + ۲s)$ و $(\bar{x} - ۳s , \bar{x} + ۳s)$ را شمارش کنید و نتیجه را به صورت درصد بیان کنید.

ج- از دامنه تغییرات برای محاسبه‌ی مقدار تقریبی انحراف معیار استفاده کنید. آیا با پاسخ قسمت الف سازگاری دارد؟

آنالیز ترکیباتی

یک سیستم مخابراتی متشکل از n آنتن کاملاً یکسان را که به ترتیب خطی به یکدیگر متصل هستند، در نظر می‌گیریم. این سیستم قادر است تا زمانی که هیچ دو آنتن متوالی معیوب نباشند همه سیگنال‌های ارسالی را دریافت کرده و یک سیستم فعال به حساب آید. حال اگر برحسب اتفاق، m تا از n آنتن معیوب باشند با چه احتمالی سیستم فعال است؟

برای مثال اگر $m=2$ و $n=4$ باشد، ۶ وضعیت متفاوت سیستم به صورت زیر خواهد بود

0	1	1	0
1	0	1	0
0	1	0	1
1	1	0	0
1	0	0	1
0	0	1	1

که ۱ نشان دهنده آنتن سالم و ۰ نشان دهنده آنتن معیوب است. بدین ترتیب فقط در سه حالت اول سیستم مخابراتی فعال است و در بقیه حالات سیستم فعال نیست، پس احتمال فعال بودن چنین سیستمی برابر $\frac{3}{6}=\frac{1}{2}$ خواهد بود. در حالت کلی برای m,n نیز می‌توان احتمال فعال بودن سیستم را با شمارش وضعیت‌هایی که سیستم فعال است نسبت به کل وضعیت‌ها محاسبه شود.

اصل اساسی شمارش: فرض کنید دو آزمایش باید اجرا شوند. اگر آزمایش ۱ به m طریق مختلف انجام شود و برای هر نتیجه آن، n طریق برای انجام آزمایش ۲ وجود داشته باشد آنگاه برای انجام تمام دو آزمایش mn نتیجه ممکن وجود خواهد داشت.

اثبات اصل اساسی:

اصل اساسی شمارش با شمارش نتایج ممکن دو آزمایش که به صورت زیر ارائه می‌شود اثبات می‌گردد

$(1,1)$, $(1,2)$ ... $(1,n)$

$(2,1)$, $(2,2)$ ... $(2,n)$

⋮

$(m,1), (m,2), \ldots, (m,n)$

بنابراین مجموعه نتایج ممکن شامل m سطر است که هر سطر دارای n عضو بوده و نتیجه ثابت می‌شود.

مثال: جامعه‌ای مرکب متشکل از ۱۰ مادر است که هر کدام ۳ فرزند دارند. اگر بخواهیم یکی از مادران و یکی از فرزندانش را به عنوان مادر و فرزند نمونه سال انتخاب کنیم، چند حالت ممکن وجود دارد؟

تعداد حالتهای ممکن ۳۰ = ۳ × ۱۰

توجه کنید که اصل اساسی شمارش در حقیقت اصل ضرب نامیده می‌شود.

تعمیم اصل اساسی ضرب: فرض کنید n عمل به صورت پی در پی انجام می‌گیرد. اگر اولین عمل را بتوان به یکی از m_1 راه متمایز انجام داد، عمل دوم را بتوان به یکی از m_2 راه متمایز انجام داد و غیره، در آن صورت n عمل مزبور با هم را به $m_1 m_2 \cdots m_n$ راه متمایز می‌توان انجام داد.

مثال: اعضای شورای دانشجویی یک دانشگاه متشکل از ۳ نفر دانشجوی سال اول، ۴ نفر دانشجوی سال دوم، ۵ نفر دانشجوی سال سوم و ۲ نفر دانشجوی سال چهارم هستند. اگر بخواهیم یک شورای فرعی ۴ نفری که دانشجویان سالهای مختلف در آن هستند را از این شورا انتخاب کنیم، چند شورای فرعی می‌توان انتخاب کرد؟

۱۲۰ = ۲ × ۵ × ۴ × ۳

مثال: تحت هر یک از سه شرط زیر، تعداد عددهای سه رقمی را که با رقم های ۵،۴،۳،۲،۰ می‌توان ساخت، تعیین کنید.

الف - بدون تکرار ارقام   ب - فرد و بدون تکرار ارقام   ج - بزرگتر از ۳۳۰ و بدون تکرار ارقام

۵	۵	۴

۱۰۰

۴	۴	۳

۴۸

بزرگتر از ۳۳۰

۱	۲	۴

۳   ۴ ۵

بزرگتر از ۴۰۰

۴۸ = ۴۰ + ۸

۲	۵	۴

بدیهی است تعداد عددهای زوج بدون تکرار برابر است با ۱۰۰ - ۴۸ = ۵۲

اصل جمع: می‌خواهیم یک عمل از n عمل متمایز را انجام دهیم. اگر عمل iام را بتوان به m_i راه متمایز انجام داد، آن‌گاه عمل مزبور را به $m_1+m_2+\dots+m_n$ راه می‌توان انجام داد.

مثال: با ارقام ۱، ۲، ۳، ۴، ۵ چند عدد بدون تکرار ارقام می‌توان نوشت به طوری که هر عدد دست‌کم سه رقم داشته باشد.

۵	۴	۳

۵	۴	۳	۲

۵	۴	۳	۲	۱

$$60+120+120=300$$

مثال: تعداد عددهای صحیح x، $1 \leq x \leq 600$، را طوری پیدا کنید که

الف- بر ۳ قابل قسمت نباشد. ب- بر ۳ یا ۵ قابل قسمت نباشد.

ج- بر ۳ یا ۵ یا ۷ قابل قسمت نباشد.

الف- $[\frac{600}{3}]=200$ تعداد اعدادی است که بر سه قابل قسمت هستند. بنابراین تعداد عددهای صحیح که بر ۳ قابل قسمت نیستند $600-200=400$ است.

ب- تعداد اعدادی که بر ۳ قابل قسمت هستند ۲۰۰ عدد بود و تعداد ارقامی که بر ۵ قابل قسمت می‌باشد $[\frac{600}{5}]=120$ است و نیز تعداد ارقامی که هم بر ۵ و هم بر ۳ قابل قسمت هستند یعنی بر عدد ۱۵ قابل قسمت هستند $[\frac{600}{15}]=40$ است بنابراین تعداد حالت‌های مساعد برای ما

$$600-200-120+40=320$$

ج- تعداد اعدادی که بر ۷ بخش‌پذیر هستند $[\frac{600}{7}]=85$ و نیز تعداد اعدادی که بر ۷ و ۵ قابل قسمت هستند $[\frac{600}{35}]=17$ و تعداد اعدادی که بر ۷ و ۳ قابل قسمت هستند $[\frac{600}{21}]=28$ و همچنین تعداد اعدادی که بر ۳ و ۵ و ۷ بخش‌پذیر هستند $[\frac{600}{105}]=5$ است. لذا تعداد عددهای صحیح که بر ۳ یا ۵ یا ۷ قابل قسمت باشند

$$600-200-120-85+40+28+17-5=275$$

مثال*: چند پلاک نمره خودرو ۷ شماره‌ای که سه شماره اول آن از حروف لاتین و چهار شماره آخر آن از ارقام ۰ تا ۹ تشکیل شده است می‌توان تهیه نمود؟

$$26 \times 26 \times 26 \times 10 \times 10 \times 10 \times 10 = 175760000$$

مثال: چند تابع را می‌توان روی n نقطه تعریف نمود، اگر هر تابع بتواند فقط مقادیر ۰ و ۱ را داشته باشد؟

اگر n نقطه را با ۱، ۲، ...، n نشان دهیم، $f(i)$ باید برای هر i $(i=1,\dots,n)$ ۰ یا ۱ باشد پس تعداد حالات ممکن توابع برابر با $2 \times 2 \times \dots \times 2 = 2^n$ است.

در مثال *: چند پلاک نمره خودرو می‌توان تهیه نمود، اگر تکرار حروف و ارقام مجاز نباشد؟

$$26 \times 25 \times 24 \times 10 \times 9 \times 8 \times 7 = 78624000$$

جایگشت

یک مجموعه مرتب n تایی از N شیء متمایز در حالتی که هیچ‌گونه تکراری بین اجزای آن نباشد را در اصطلاح یک جایگشت n تایی از N شیء مزبور می‌نامیم و تعداد این جایگشت‌ها را با $P(N,n)$ نشان می‌دهیم

$$P(N,n) = \frac{N!}{(N-n)!}$$

حالت خاص: $P(N,N) = N!$ یعنی این است که N شیء را به $N!$ راه می‌توان مرتب کرد.

مثال: الف - چند کلمه ۹ حرفی بدون تکرار می‌توان از حروف کلمه BOULEVARD ساخت؟

ب - در چه تعداد از این کلمه‌ها حرف‌های O و U پشت سر هم می‌آیند؟

ج - در چه تعداد از این کلمه‌ها حرف‌های O و U پشت سر هم نیستند؟

الف - $P(9,9) = 9! = 362880$

ب - $2 \times 8! = 8! + 8! = P(8,8) + P(8,8) = 80640$

$$362880 - 80640 = 282240$$ ج -

مثال: ده صندلی در یک ردیف قرار گرفته اند و ۲ نفر می‌خواهند روی این صندلی‌ها بنشینند.

الف - احتمال اینکه این دو نفر پهلوی هم قرار گیرند را بیابید.

ب - احتمال اینکه بین این دو نفر یک صندلی قرار بگیرد را بیابید.

الف - اگر A پیشامد قرار گرفتن دو نفر پهلوی هم باشد چون به ۹ طریق می‌توان دو صندلی پهلوی هم را انتخاب کرد پس

$$n(S) = P(10,2) = \frac{10!}{8!} = 90$$

$$n(A) = 9 \times 2! = 18 \qquad P(A) = \frac{18}{90} = \frac{1}{5}$$

ب -

$$P(B) = \frac{8 \times 2!}{90} = \frac{16}{90} = \frac{8}{45}$$

مثال: شخصی ۱۰ کتاب دارد و می‌خواهد آنها را در قفسه کتابخانه خود قرار دهد. ۱۰ کتاب، به تعداد ۴ کتاب ریاضی، ۳ کتاب شیمی، ۲ کتاب تاریخ و یک کتاب زبان است. اگر این شخص بخواهد کتابهای با موضوع یکسان کنار هم قرار گیرد، چند ترتیب ممکن برای قرار دادن کتابها وجود دارد؟

تعداد $1!2!3!4!$ حالت ممکن وجود دارد که ابتدا کتابهای ریاضی، بعد کتابهای شیمی، سپس کتابهای تاریخ و در انتها کتاب زبان قرار گیرد. اما ترتیب موضوعی کتابها خود نیز به $4!$ حالت امکان پذیر است. بنابراین پاسخ مورد نظر برابر با $4!(4!3!2!1!) = 6912$ است.

مثال: چند ترتیب متمایز از حروف کلمه PEPPER می‌توان تهیه نمود؟

نخست توجه کنید که $6!$ جایگشت برای حروف $P_1E_1P_2P_3E_2R$ وجود دارد که ۳ حرف P و ۲ حرف E از یکدیگر متمایز باشند. بنابراین اگر یکی از جایگشتها را مثلاً $P_1P_2E_1P_3E_2R$ در نظر بگیریم و حروف P را بین خودشان و حروف E را بین خودشان

جابجا کنیم باز هم در نتیجه همچنان PPEPER است. بنابراین تمام !۳!۲ جایگشت زیر به ترتیب دارند همان ترتیب PPEPER است.

$P_1 P_2 E_1 P_3 E_2 R$ $\qquad$ $P_1 P_2 E_2 P_3 E_1 R$

$P_1 P_3 E_1 P_2 E_2 R$ $\qquad$ $P_1 P_3 E_2 P_2 E_1 R$

$P_2 P_1 E_1 P_3 E_2 R$ $\qquad$ $P_2 P_1 E_2 P_3 E_1 R$

$P_2 P_3 E_1 P_1 E_2 R$ $\qquad$ $P_2 P_3 E_2 P_1 E_1 R$

$P_3 P_1 E_1 P_2 E_2 R$ $\qquad$ $P_3 P_1 E_2 P_2 E_1 R$

$P_3 P_2 E_1 P_1 E_2 R$ $\qquad$ $P_3 P_2 E_2 P_1 E_1 R$

بنابراین تعداد جایگشت‌های ممکن حروف کلمه PEPPER برابر $\frac{6!}{3!\,2!}=60$ است.

تذکر: به طور کلی تعداد جایگشت‌های n شیء که n_1 تا از آن‌ها مشابه هم، n_2 تا از آن‌ها مشابه هم، ... و n_r تا از آن‌ها مشابه هم هستند برابر است با

$$\frac{n!}{n_1!\,n_2!\cdots n_r!}$$

مثال: در یک مسابقه از ۱۰ بازیکن، ۴ نفر روسی، ۳ نفر آمریکایی، ۲ نفر انگلیسی و یک نفر برزیلی هستند. اگر نتیجه مسابقه تنها ملیت بازیکنان را به ترتیبی که برنده شده‌اند اعلام کند، چند نتیجه ممکن وجود دارد؟

$$\frac{10}{4!\,3!\,2!\,1!}=12600$$

مثال: چند علامت مختلف را که هرکدام شامل ۹ پرچم قرار گرفته در یک خط هستند به وسیله ۴ پرچم سفید، ۳ پرچم قرمز و ۲ پرچم آبی می‌توان تهیه شود؟ پرچم‌های هم رنگ یکسان هستند.

$$\frac{9!}{4!\,3!\,2!}=1260$$

ترکیب: یک نمونه نامرتب r تایی از n شیء متمایز درحالتی که نمونه گیری بدون جایگذاری است را در اصطلاح یک ترکیب r تایی از n شیء مزبور می نامیم و تعداد ترکیب های ممکن را با $C(n,r)$ یا $\binom{n}{r}$ نشان می دهیم

$$C(N,n) = \frac{n!}{(n-r)!\,r!}$$

به طور کلی چون $n(n-1)\cdots(n-r+1)$ برابر است با تعداد ترتیب r عضوی از مجموعه n عضوی و چون هر گروه r عضوی به تعداد $r!$ ترتیب دارد، بنابراین تعداد حالتهای متفاوت انتخاب گروه r عضوی از یک مجموعه n عضوی برابر است با

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!}$$

مثال: می خواهیم از میان یک گروه ۲۰ نفری یک شورای سه نفره تشکیل دهیم. به چند حالت این کار امکان پذیر است؟

$$\binom{20}{3} = \frac{20\times 19\times 18}{3\times 2\times 1} = 1140$$

مثال: از یک گروه متشکل از ۵ زن و ۷ مرد، چند شورای مختلف ۵ عضوی شامل ۲ زن و ۳ مرد می توان انتخاب نمود؟

چون $\binom{5}{2}$ تعداد حالتهای ممکن انتخاب گروه ۲ نفری از زنها و $\binom{7}{3}$ تعداد حالتهای ممکن انتخاب گروه ۳ نفری از مردهاست، بنابراین براساس اصل اساسی شمارش تعداد شوراهای ۲ زن و ۳ مرد برابر است با

$$\binom{5}{2}\binom{7}{3} = \frac{5\times 4}{2\times 1}\times\frac{7\times 6\times 5}{3\times 2\times 1} = 350$$

حال اگر دو نفر از مردها نخواهند باهم انتخاب شوند، تعداد حالتهایی که گروه ۳ نفری مردها بدون آن دو می توانند انتخاب شوند برابر با $\binom{2}{0}\binom{5}{3}$ است. از طرف دیگر تعداد حالتهایی که در گروه ۳ نفری یکی از آن دو نفر انتخاب شوند برابر با $\binom{2}{1}\binom{5}{2}$ است.

پس تعداد کل گروههای سه نفری از مردها که آن دو نفر باهم نیستند برابر است با

$$\binom{2}{0}\binom{5}{3} + \binom{2}{1}\binom{5}{2} = 30$$

بنابراین تعداد کل حالتهای ممکن انتخاب شده برای ۵ نفره برابر $\binom{5}{2} = ۳۰$ برابر ۳۰ است.

مثال: مجموعه‌ای از n آنتن را درنظر بگیرید که m تای آنها معیوب و بقیه سالم می‌باشد. فرض کنید آنتنهای سالم و معیوب غیرقابل تشخیص باشند. چند ترتیب خطی از آنتنها وجود دارد به طوری که هیچ دو آنتن معیوب به طور متوالی قرار نگیرند؟

حل: فرض کنید که $n-m$ آنتن سالم را در یک ردیف قرار دهیم. حال برای اینکه هیچ دو آنتن معیوب پهلوی هم نباشند باید آنها را در فاصله‌های بین آنتنهای سالم قرار داد. به طوری که در هر فاصله حداکثر یک آنتن معیوب قرار گیرد. یعنی از $n-m+1$ موقعیت ممکن باید m تا را انتخاب نموده و آنتنهای معیوب را در آنجا قرار دهیم.

$$- \mid - \mid - \mid \cdots - \mid - \mid -$$

| نشان دهنده آنتن سالم و $-$ جایی که حداکثر یک آنتن معیوب قرار گیرد.

بنابراین به تعداد $\binom{n-m+1}{m}$ حالت ممکن وجود دارد که حداقل یک آنتن سالم بین دو آنتن معیوب قرار گرفته باشد.

تذکر:
$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

فرض کنید مجموعه A شامل n عضو است مثلاً $A = \{1, 2, \ldots, n\}$ تعداد انتخابهای r عضو از میان n عضو $\binom{n}{r}$ است. برای نشان دادن درستی رابطه فوق یک عضو خاص مثلاً ۱ را درنظر می‌گیریم. به طور انتخاب r عضو از میان n عضو این امکان وجود دارد که یا ۱ متعلق به این مجموعه r عضوی باشد و یا آنکه ۱ به این مجموعه تعلق نداشته باشد. اگر ۱ متعلق به مجموعه r عضوی باشد بدیهی است $r-1$ عضو باقیمانده را می‌توان به انتخاب $r-1$ عضو از میان $n-1$ عضو انتخاب کرد. حال اگر

۱ متعلق به مجموعه r عضوی نباشد آنگاه باید r عضو را از بین n عضو منهای عدد ۱ یعنی n-1 عضو انتخاب نمود. بنابراین تعداد طرقی که می‌توان r عضو را انتخاب نمود $\binom{n-1}{r}+\binom{n-1}{r-1}$ است.

راه حل دوم:

$$\binom{n-1}{r-1}+\binom{n-1}{r}=\frac{(n-1)!}{(r-1)!\,(n-r)!}+\frac{(n-1)!}{r!\,(n-1-r)!}$$

$$=\frac{r\,(n-1)!+(n-r)\,(n-1)!}{r!\,(n-r)!}=\frac{n\,(n-1)!}{r!\,(n-r)!}=\binom{n}{r}$$

مقدار $\binom{n}{r}$ را اغلب ضریب دو جمله‌ای می‌نامند، زیرا در قضیه دو جمله‌ای نقش ضریب را ایفا می‌کند.

قضیه دوجمله‌ای: $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$

اثبات:

$$n=1 \Rightarrow (x+y)=\sum_{k=0}^{1}\binom{1}{k}x^k y^{1-k}=\binom{1}{0}x^0y^1+\binom{1}{1}x^1y^0=x+y$$

حال فرض کنیم رابطه برای n-1 برقرار است، پس

$$(x+y)^n=(x+y)(x+y)^{n-1}=(x+y)\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-1-k}$$

$$=\sum_{k=0}^{n-1}\binom{n-1}{k}x^{k+1}y^{n-1-k}+\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k}$$

حال اگر $i=k+1$ بگیریم آنگاه

$$(x+y)^n=\sum_{i=1}^{n}\binom{n-1}{i-1}x^i y^{n-i}+\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k}$$

$$=\sum_{k=1}^{n}\binom{n-1}{k-1}x^k y^{n-k}+\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k}$$

$$=\sum_{k=1}^{n-1}\binom{n-1}{k-1}x^k y^{n-k}+x^n+y^n+\sum_{k=1}^{n-1}\binom{n-1}{k}x^k y^{n-k}$$

$$=\sum_{k=1}^{n-1}\left[\binom{n-1}{k-1}+\binom{n-1}{k}\right]x^k y^{n-k}+x^n+y^n$$

$$=\sum_{k=1}^{n-1}\binom{n}{k}x^k y^{n-k}+x^n+y^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$

اثبات ترکیبی قضیه دوجمله‌ای: حاصل‌ضرب زیر را در نظر می‌گیریم

$$(x_1+y_1)(x_2+y_2)\cdots(x_n+y_n)$$

بسط حاصل‌ضرب فوق به صورت جمع 2^n عضو ظاهر می‌شود که هر عضو حاصل‌ضرب n عامل بوده و به علاوه هر یک از 2^n عضو دارای عامل x_i یا y_i برای $i=1,\ldots,n$ است.

برای مثال

$$(x_1+y_1)(x_2+y_2)=x_1x_2+x_1y_2+x_2y_1+y_1y_2$$

حال چند عضو از 2^n عضو دارای k عامل x_i و $(n-k)$ عامل y_i می‌باشند؟

نظر به اینکه هر عضو که از k تا x_i و $(n-k)$ تا y_i تشکیل شده در ارتباط با انتخاب یک گروه k تایی از n مقدار $x_1, x_2, \ldots, x_n$ است، پس جمعاً $\binom{n}{k}$ از چنین جملات وجود دارد. حال با در نظر گرفتن $x_i=x$ و $y_i=y$ برای $i=1,2,\ldots,n$ داریم

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$

مثال: عبارت $(x+y)^3$ را بسط دهید.

$$(x+y)^3=\binom{3}{0}x^0y^3+\binom{3}{1}x^1y^2+\binom{3}{2}x^2y^1+\binom{3}{3}x^3y^0=y^3+3xy^2+3x^2y+x^3$$

مثال: چند زیرمجموعه از یک مجموعه n عضوی وجود دارد؟

$$\sum_{k=0}^{n}\binom{n}{k}=(1+1)^n=2^n$$

راه حل دوم: این نتیجه را می‌توان با متناظر کردن صفر و یک به هر عضو مجموعه براساس شیوه رایج تک به تک نیز به دست آورد. اگر برای هر تناظر، زیرمجموعه‌ای از عناصری که عدد ۱ به آنها متناظر کرده‌ایم تهیه کنیم، چون تعداد تناظر برابر با 2^n است نتیجه مطلوب حاصل می‌شود.

توجه کنید که یکی از زیرمجموعه‌ها، مجموعه‌ای است که همه عناصر آن صفر هستند. یعنی مجموعه تهی و بنابراین تعداد زیرمجموعه‌هایی که حداقل یک عضو دارند برابر 2^n-1 است.

ضرایب چند جمله‌ای

مجموعه‌ای متشکل از n عضو متفاوت را باید به r گروه مجزا با اندازه‌های به ترتیب n_1، n_2، ...، و n_r تقسیم کنیم به طوری که $\sum_{i=1}^{r} n_i = n$ به چند روش این کار ممکن است؟

حل: $\binom{n}{n_1}$ انتخاب ممکن برای گروه ۱ وجود دارد. برای هر انتخاب گروه ۱ به تعداد $\binom{n-n_1}{n_2}$ انتخاب ممکن برای گروه ۲ وجود خواهد داشت و به همین ترتیب برای گروه ۳ تعداد انتخاب‌های ممکن برابر $\binom{n-n_1-n_2}{n_3}$ است و ... بنابراین طبق اصل ضرب (طبق تعمیم اصل اساسی شمارش) تعداد حالتهای ممکن برابر است با

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\times\cdots\times\binom{n-n_1-\cdots-n_{r-1}}{n_r}$$

$$=\frac{n!}{(n-n_1)!\,n_1!}\times\frac{(n-n_1)!}{(n-n_1-n_2)!\,n_2!}\times\cdots\times\frac{(n-n_1-\cdots-n_{r-1})!}{0!\,n_r!}$$

$$=\frac{n!}{n_1!\,n_2!\cdots n_r!}$$

نمادگذاری:

اگر $n_1+n_2+\cdots+n_r=n$ باشد $\binom{n}{n_1,n_2,\ldots,n_r}$ را به صورت زیر تعریف می‌کنیم

$$\binom{n}{n_1,n_2,\ldots,n_r}=\frac{n!}{n_1!\,n_2!\cdots n_r!}$$

بنابراین $\binom{n}{n_1,\ldots,n_r}$ نشان دهنده تعداد طرق تقسیم n شیء مختلف به r گروه متمایز است به طوری که هر گروه به ترتیب دارای n_1، n_2، ...، n_r عضو باشد.

مثال: یک مرکز پلیس در شهری کوچک دارای ۱۰ افسر پلیس است. اگر برنامه مرکز این باشد که ۵ افسر وظیفه نظارت بر خیابانها، ۲ افسر پاسداری از مرکز و ۳ افسر به صورت ذخیره باشند، به چند حالت ممکن می‌توان ۱۰ افسر را به ۳ گروه تقسیم نمود؟

$$\frac{10!}{5!\,2!\,3!} = 2520$$

مثال: ۱۰ ورزشکار را باید به دو تیم A و B هرکدام با ۵ عضو تقسیم شوند. تیم A در یک گروه از مسابقات و تیم B در گروه دیگر مسابقات شرکت می‌کنند. چند تقسیم‌بندی مختلف امکان‌پذیر است؟

حل: $$\frac{10!}{5!\,5!} = 252$$

مثال: برای انجام یک بازی بسکتبال ۱۰ ورزشکار را که در زمین بازی هستند به دو تیم ۵ تایی تقسیم می‌کنیم. چند تقسیم‌بندی مختلف امکان‌پذیر است؟

دقت شود در این مثال برخلاف مثال قبل ترتیب دو تیم اهمیت ندارد یعنی دو تیم متفاوت A و B وجود ندارد و فقط هدف تقسیم‌بندی ۱۰ نفر به دو گروه ۵ تایی است:

$$\frac{\dfrac{10!}{5!\,5!}}{2!} = 126$$

قضیه چندجمله‌ای: برای تمامی مقادیر $x_1, \dots, x_k$ و هر عدد صحیح مثبت n،

$$(x_1 + x_2 + \dots + x_k)^n = \sum_{\substack{(n_1, \dots, n_k) \\ n_1 + n_2 + \dots + n_k = n}} \binom{n}{n_1, n_2, \dots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$$

که در آن جمع روی مقادیر صحیح غیرمنفی بردارهای $(n_1, n_2, \dots, n_k)$ است، به طوری که $n_1 + n_2 + \dots + n_k = n$.

اثبات: برای اثبات از روش استقرا استفاده می‌کنیم. چون ما قبلاً قضیه دوجمله‌ای را اثبات کرده‌ایم، می‌دانیم که نتیجه برای قضیه چندجمله‌ای برای هر n اگر $k=2$ باشد صحیح است. ما ابتدا از استقرا (؟) می‌کنیم اما این بار به جای n از k استفاده می‌کنیم. برای $k=2$ قبلاً می‌دانیم که نتیجه درست است. فرض کنید که نتیجه برای هر $k \le k_0$ و برای هر n صحیح است.

برای $k = k_0+1$ و n دلخواه، باید نشان دهیم که

$$(x_1+x_2+\cdots+x_{k_0+1})^n = \sum_{(n_1,\dots,n_{k_0+1})} \binom{n}{n_1,\dots,n_{k_0+1}} x_1^{n_1}\cdots x_{k_0+1}^{n_{k_0+1}} \qquad (1)$$

$$n_1+\cdots+n_{k_0+1}=n$$

فرض کنید $y_i = x_i$ برای $i=1,\dots,k_0-1$ و فرض کنید $y_0 = x_{k_0}+x_{k_0+1}$.

آنگاه داریم

$$(x_1+\cdots+x_{k_0+1})^n = (y_1+\cdots+y_{k_0})^n$$

چون فرض کرده‌ایم که قضیه برای $k=k_0$ صحیح است، می‌دانیم که

$$(y_1+\cdots+y_{k_0})^n = \sum \binom{n}{m_1,\dots,m_{k_0}} y_1^{m_1}\cdots y_{k_0}^{m_{k_0}} \qquad (2)$$

که در آن مجموع روی تمام $m_1,\dots,m_{k_0}$ ها است به طوری که $m_1+\cdots+m_{k_0}=n$.

در سمت راست رابطه (2) $x_{k_0}+x_{k_0+1}$ را به جای y_{k_0} جایگزین کنید و سپس قضیه دوجمله‌ای را به کار برید.

$$\sum \binom{n}{m_1,\dots,m_{k_0}} y_1^{m_1}\cdots y_{k_0-1}^{m_{k_0-1}}$$

$$\times \sum_{i=0}^{m_{k_0}} \binom{m_{k_0}}{i} x_{k_0}^{i}\, x_{k_0+1}^{m_{k_0}-i} \qquad (3)$$

در (3) فرض کنید $n_i = m_i$ برای $i=1,\dots,k_0-1$ و فرض کنید $n_{k_0}=i$ و فرض کنید $n_{k_0+1} = m_{k_0}-i$. آنگاه در مجموع در (3) $n_1+\cdots+n_{k_0+1}=n$ اگر و تنها اگر $m_1+\cdots+m_{k_0}=n$. همچنین توجه کنید که

$$\binom{n}{m_1,\dots,m_{k_0}}\binom{m_{k_0}}{i} = \binom{n}{n_1,n_2,\dots,n_{k_0+1}}$$

بنابراین رابطه (3) می‌شود

$$\sum \binom{n}{n_1,\dots,n_{k_0+1}} x_1^{n_1}\cdots x_{k_0+1}^{n_{k_0+1}}$$

که در آن این آخرین مجموع روی تمام $n_1,\dots,n_{k_0+1}$ است به طوری که $n_1+\cdots+n_{k_0+1}=n$.

مقدار $\binom{n}{n_1,\dots,n_k}$ به عنوان ضریب چند جمله‌ای شناخته می‌شود.

مثال: عبارت $(x_1+x_2+x_3)^2$ را بسط دهید.

$$(x_1+x_2+x_3)^2 = \binom{2}{2,0,0}x_1^2x_2^0x_3^0 + \binom{2}{0,2,0}x_1^0x_2^2x_3^0$$
$$+\binom{2}{0,0,2}x_1^0x_2^0x_3^2 + \binom{2}{1,1,0}x_1^1x_2^1x_3^0 + \binom{2}{0,1,1}x_1^0x_2^1x_3^1$$
$$+\binom{2}{1,0,1}x_1^1x_2^0x_3^1 = x_1^2+x_2^2+x_3^2+2x_1x_2+2x_1x_3+2x_2x_3$$

تعداد جواب‌های صحیح معادلات

تعداد نتایج ممکن توزیع n توپ متمایز در r ظرف متمایز برابر r^n است. زیرا هر توپ می‌تواند در هر یک از r ظرف قرار گیرد.

برای بدست آوردن تعداد نتایج ممکن توزیع n توپ یکسان در r ظرف متمایز به طریق زیر عمل می‌کنیم:

چون توپ‌ها یکسان هستند، بنابراین نتیجه آزمایش توزیع n توپ به r ظرف را می‌توان به صورت بردار $(x_1,x_2,\dots,x_r)$ بیان نمود. x_i نشان دهنده تعداد توپ‌هایی است که در ظرف iام قرار دارند. بنابراین مسئله به صورت تعیین تعداد اعداد صحیح غیرمنفی عناصر $(x_1,\dots,x_r)$ برداری که در آن $x_1+x_2+\dots+x_r=n$ ساده می‌شود.

برای محاسبه ابتدا فرض می‌کنیم که همه عناصر اعداد صحیح مثبت هستند. تصور می‌کنیم که n توپ یکسان را در یک ردیف قرار داده و می‌خواهیم آنها را به r گروه غیرتهی تقسیم کنیم. برای این منظور می‌توانیم از $n-1$ فاصله بین توپ‌ها $r-1$ تا را انتخاب نموده و آنها را نقاط تقسیم توپ‌ها در ظرف‌ها در نظر بگیریم. برای مثال اگر $n=8$ و $r=3$ باشد و ۲ تقسیم به صورت زیر در نظر بگیریم

$$\circ\circ\circ\,|\,\circ\circ\circ\,|\,\circ\circ$$

آنگاه عناصر بردار برابر با $x_1 = 3$ ، $x_2 = 3$ و $x_3 = 2$ است. تعداد حالت‌های تعیین انتخاب برابر با $\binom{n-1}{r-1}$ است. $0 \wedge 0 \wedge 0 \wedge \cdots \wedge 0 \wedge 0$

گزاره تعداد $\binom{n-1}{r-1}$ بردار متمایز r عنصری $(x_1, x_2, \ldots, x_r)$ با عناصر صحیح مثبت x_i ، $x_i > 0$ برای $i = 1, 2, \ldots, r$ و با شرط $x_1 + x_2 + \cdots + x_r = n$ وجود دارد.

نکته: تعداد جواب‌های غیرمنفی معادله $x_1 + x_2 + \cdots + x_r = n$ ، $\forall i \; x_i \geq 0$ مانند تعداد جواب‌های مثبت معادله

$$y_1 + y_2 + \cdots + y_r = n + r \qquad \forall i \; y_i \geq 1$$

است. در حقیقت $i = 1, \ldots, r$ ، $y_i = x_i + 1$.

تعداد $\binom{n+r-1}{r-1}$ بردار متمایز r عنصری $(x_1, \ldots, x_r)$ با عناصر صحیح غیرمنفی x_i با شرط $x_1 + x_2 + \cdots + x_r = n$ وجود دارد.

مثال: چند جواب متمایز غیرمنفی صحیح برای معادله $x_1 + x_2 = 3$ وجود دارد؟

حل: $\binom{3+2-1}{2-1} = 4$ جواب وجود دارد که عبارتند از

$(3, 0)$ ، $(2, 1)$ ، $(1, 2)$ ، $(0, 3)$

مثال: سرمایه‌گذاری مایل است ۲۰ میلیون تومان را در ۴ زمینه متفاوت سرمایه‌گذاری کند و در هر زمینه باید مضربی از میلیون تومان سرمایه‌گذاری شود. اگر وی بخواهد همه ۲۰ میلیون تومان را سرمایه‌گذاری کند چند روش مختلف وجود دارد؟ چنانچه لازم نباشد همه سرمایه را مصرف کند چند روش وجود دارد؟

$$x_1 + x_2 + x_3 + x_4 = 20 \qquad x_i \geq 0$$

x_i : مقدار سرمایه‌گذاری در زمینه i (با واحد میلیون تومان)
$i = 1, 2, 3, 4$

$$\binom{n+r-1}{r-1} = \binom{20+4-1}{4-1} = \binom{23}{3}$$

اگر لازم نباشد که همه سرمایه مصرف شود، فرض می‌کنیم که مقداری x میلیون تومان آن باقی بماند.

$$x_1+x_2+x_3+x_4+x_5=20 \qquad \forall i \ \ x_i\geq 0 \quad i=1,2,3,4,5$$

$$\binom{20+5-1}{5-1}=\binom{24}{4}=10626$$

مثال: چند جمله در بسط چندجمله‌ای $(x_1+x_2+\dots+x_r)^n$ وجود دارد.

$$(x_1+x_2+\dots+x_r)^n=\sum_{\substack{(n_1,\dots,n_r)\\ n_1+n_2+\dots+n_r=n}}\binom{n}{n_1,\dots,n_r}x_1^{n_1}\cdots x_r^{n_r}$$

و به ازای همه $n_i\geq 0$. بنابراین تعداد جملات برابر $\binom{n+r-1}{r-1}$ است.

مثال: مجدداً به مثال مجموعه‌ای از n جزء صف، که m تای آن معیوب و بقیه $n-m$ جزء سالم بودند مراجعه می‌کنیم. در این مثال هدف تعیین تعداد ترتیب خطی اجزاء بود که هیچ دو جزء متوالی معیوب نباشند. برای حل ابتدا اجزاء معیوب را در یک ردیف قرار می‌دهیم و بین آن‌ها حداقل یک جزء سالم قرار می‌دهیم.

$$x_1 \bigcirc x_2 \bigcirc x_3 \bigcirc x_4 \bigcirc \cdots \bigcirc x_m \bigcirc x_{m+1}$$

$$x_1\geq 0 \quad x_2>0 \quad x_3>0 \quad x_4>0 \quad x_m>0 \quad x_{m+1}\geq 0$$

$$x_1,x_{m+1}\geq 0 \qquad \forall i=2,\dots,m \quad x_i\geq 1$$

$$x_1+x_2+x_3+\dots+x_m+x_{m+1}=n-m$$

$$x_i=y_i \qquad \forall i=2,\dots,m$$

$$1+x_1=y_1 \qquad y_{m+1}=x_{m+1}+1$$

$$\Rightarrow y_1+y_2+y_3+\dots+y_m+y_{m+1}=n-m+2 \qquad \forall i \ \ y_i\geq 1$$

$$\binom{n-m+2-1}{m+1-1}=\binom{n-m+1}{m}$$

حال فرض کنید پاسخ را در حالتی بخواهیم که هر دو جزء معیوب حداقل به وسیله دو جزء سالم جدا شده باشند.

$$x_1 + x_2 + \cdots + x_m + x_{m+1} = n-m$$

$$i = 2, \dots m \quad x_i \geqslant 2 \qquad x_1 \geqslant 0 \qquad x_{m+1} \geqslant 0$$

$$y_1 = x_1 \qquad y_i = x_i - 2 \quad i = 2, \dots, m \qquad y_{m+1} = x_{m+1} \qquad \forall i \quad y_i \geqslant 0$$

$$y_1 + y_2 + \cdots + y_m + y_{m+1} = n - m - 2(m-1) = n - 3m + 2$$

بنابراین تعداد جوابهای غیرمنفی معادله $y_1 + \cdots + y_{m+1} = n - 3m + 2$ پاسخ مطلوب است

$$\binom{n-3m+2+m}{m+1-1} = \binom{n-2m+2}{m}$$

به نام خدا

مفاهیم مقدماتی احتمال

آزمایش تصادفی : هر آزمایش (پدیده یا تجربه‌ای) که نتایج انجام آن از پیش به طور قطع قابل پیش بینی نباشد ولی همه نتایج ممکن آن معلوم باشند را یک آزمایش تصادفی گوییم . چون در احتمال اصولاً با آزمایش‌های تصادفی سروکار داریم غالباً آن را به طور خلاصه آزمایش می‌گوییم .

بدیهی است آزمایش‌های غیر تصادفی یا قطعی آنهایی هستند که نتایج انجام آنها از پیش به طور یقین معلوم باشد .

رویداد یا برآمد : به هر نتیجه ممکن یک آزمایش تصادفی یک رویداد یا برآمد آن می‌گوییم .

فضای نمونه : به مجموعه همه رویدادهای ممکن یک آزمایش تصادفی فضای نمونه آن می‌گوییم و آن را با نماد S نشان می‌دهیم .

پیشامد : به هر مجموعه‌ای از رویدادهای یک آزمایش تصادفی یک پیشامد آن آزمایش گوییم . پیشامدها را غالباً با حروف بزرگ لاتین A ، B ، C ، ... نشان می‌دهیم . لذا A ، B ، C ، ... زیر مجموعه‌های S هستند . پیشامدها وقتی رخ می‌دهند که در انجام آزمایش یکی از رویدادهای متعلق به آنها اتفاق افتد .

تعریف : دو پیشامد A و B را ناسازگار یا جدا از هم گوییم هرگاه نتوانند با هم رخ دهند . به عبارت دیگر ، هرگاه $A \cap B = \phi$.

مثال : پرتاب یک سکه یک آزمایش تصادفی است با $S = \{T, H\}$ که در آن H ، نشان دهندهٔ «روی سکه» (اصطلاحاً شیر) ، و T نشان دهنده «پشت سکه» (اصطلاحاً خط) ، رویدادهای آزمایش هستند . $A = \{T\}$ یک پیشامد این آزمایش است .

تذکر: T یک رویداد است ولی $\{T\}$ یک پیشامد محسوب می‌شود.

مثال: پرتاب یک شش وجهی منتظم که روی وجوه آن اعداد ۱ تا ۶ نوشته شده است (که اصطلاحاً تاس نامیده می‌شود) یک آزمایش تصادفی است با فضای نمونه $S=\{1,2,3,4,5,6\}$.

"زوج بودن شماره تاس" یک پیشامد آن محسوب می‌شود که عبارتست از:

$$A=\{\text{زوج بودن شماره}\}=\{2,4,6\}.$$

$$B=\{\text{مضرب ۳ بودن شماره}\}=\{3,6\}$$

پیشامد دیگری از این آزمایش است.

$$A\cap B=\{6\}$$ A و B ناسازگار نیستند

مثال: پرتاب همزمان یک سکه و یک تاس و مشاهده نتایج آنها یک آزمایش تصادفی است، با فضای نمونه

$$S=\{(1,H),(2,H),(3,H),(4,H),(5,H),(6,H),$$
$$(1,T),(2,T),(3,T),(4,T),(5,T),(6,T)\}$$

زوج بودن شماره تاس و خط بودن سکه یک پیشامد این آزمایش است که دارای ۳ رویداد می‌باشد.

$$A=\{(2,T),(4,T),(6,T)\}$$

مثال: فاصله مرز تا محل توقف توپی که او پرتاب می‌کند یک فاصله تصادفی است با فضای نمونه $S=\{x:0<x<l\}$ که در آن l ماکسیمم فاصله‌ای است که توپ می‌تواند برود.

$$A=\{\text{حداقل ۱۰۰ متر فاصله}\}=\{x:100\leq x<l\}$$

یک پیشامد این آزمایش است.

مثال: از خط تولید یک کارخانه ۳ محصول را به طور تصادفی انتخاب می‌کنیم. این محصولات ممکن است خراب یا سالم باشند.

الف - اگر خراب بودن محصول را با D و سالم بودن آن را با N نمایش دهیم آنگاه فضای نمونه مورد نظر عبارت است از :

$$S_1 = \{NNN, NND, NDN, DNN, DDN, DND, NDD, DDD\}$$

ب - اگر به تعداد قطعات خراب در بین ۳ قطعه انتخابی توجه کنیم آنگاه فضای نمونه عبارت است از :

$$S_2 = \{0, 1, 2, 3\}$$

بنابراین در یک آزمایش تصادفی ممکن است بیش از یک فضای نمونه داشته باشیم وجنبه مورد نظر از آزمایش تصادفی است که فضای نمونه را تعیین می کند. همچنین باتوجه به مثالهای فوق می توان فضاهای نمونه را به طور کلی به دو گروه زیر تقسیم کرد :

۱ - فضای نمونه گسسته

الف - فضای نمونه متناهی که تعداد اعضای آن متناهی است .

ب - فضای نمونه نامتناهی شمارش پذیر : که یک مجموعه نامتناهی اما شمارش پذیر است .

۲ - فضای نمونه پیوسته که اعضای آن به صورت یک فاصله از اعداد حقیقی یا یک سطح در فضای دو بعدی یا ... است مانند مثال فاصله فرد تا محل توقف توپ .

وقوع (رخداد) یک پیشامد : گوئیم پیشامد A به وقوع پیوسته است ، هرگاه نتیجه آزمایش تصادفی منجر به مشاهده عضوی از پیشامد A گردد . برای مثال در آزمایش تصادفی پرتاب یک تاس اگر شماره رویه مشاهده شده مد نظر ما باشد و پیشامد A پیشامد زوج بودن شماره خال مشاهده شده یعنی $A = \{2, 4, 6\}$ و B پیشامد فرد بودن شماره رویه مشاهده شده باشد یعنی $B = \{1, 3, 5\}$ آنگاه اگر شماره ۴ شماره حاصل از پرتاب باشد ، پیشامد A روی داده است . اگر شماره حاصل از پرتاب

عدد ۵ باشد گوییم پیشامد B رخ داده است . اگر پیشامد $C = \{1, 2, 5\}$

و نتیجه حاصل از پرتاب عدد ۲ باشد آنگاه هم پیشامد A و هم پیشامد C روی داده

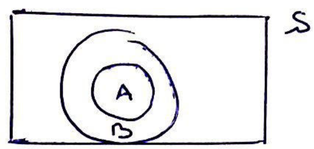
است .

اعمال روی پیشامدها

چون پیشامدها زیرمجموعه‌هایی از فضای نمونه هستند پس می‌توان مانند مجموعه‌ها اعمال

جبری را روی آن‌ها انجام داد . در این حالت فضای نمونه مجموعه مرجع می‌باشد و

در نمودار ون آن را با مستطیل و پیشامدها را با دایره نشان می‌دهند .



$A \subset B$

الف - زیرپیشامد : پیشامد A را زیرپیشامد B گوییم هرگاه وقوع A وقوع B را نتیجه

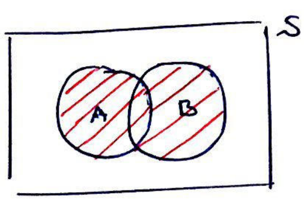
دهد و آن را با نماد $A \subset B$ نمایش می‌دهیم .

ب - دو پیشامد مساوی : دو پیشامد A و B را مساوی گوییم هرگاه وقوع یکی وقوع دیگری را

نتیجه دهد : یعنی $A = B \Leftrightarrow (A \subset B \text{ و } B \subset A)$

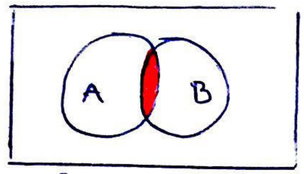
پ - اجتماع دو پیشامد : پیشامد $A \cup B = \{x \mid x \in A \text{ یا } x \in B\}$ را اجتماع دو پیشامد A و B

گوییم و وقوع $A \cup B$ به معنای وقوع حداقل یکی از دو پیشامد A یا B است

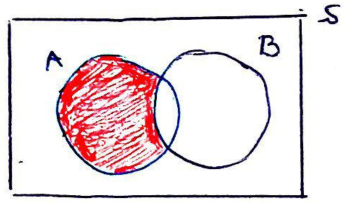


ت - اشتراک دو پیشامد : پیشامد $A \cap B = \{x \mid x \in A , x \in B\}$ را اشتراک دو پیشامد A

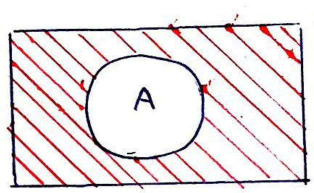
و B گوییم و وقوع $A \cap B$ به معنای رخداد هم‌زمان دو پیشامد A و B است .



ب- تفاضل دو پیشامد: پیشامد $A-B=\{x \mid x\in A ,\ x\notin B\}$ را تفاضل پیشامد B از A گوییم و وقوع $A-B$ به معنای وقوع "فقط A و نه B" است.



ج- متمم یک پیشامد: پیشامد $A^c=\{x \mid x\in S ,\ x\notin A\}$ را متمم پیشامد A گوییم و وقوع A^c به معنای عدم وقوع پیشامد A است.



اگر $E_1 , E_2 , \dots$ پیشامد باشند آنگاه اجتماع این پیشامدها که به صورت $\bigcup_{n=1}^{\infty} E_n$ نشان داده می‌شود عبارت است از پیشامدی که شامل همه نقاطی باشد که در E_n برای حداقل یک مقدار n ($n=1,2,\dots$) وجود داشته باشد. اشتراک E_n ها به صورت $\bigcap_{n=1}^{\infty} E_n$ نشان داده می‌شود و برابر با پیشامدی است که نقاط آن در همه پیشامدهای E_n ($n=1,2,\dots$) وجود داشته باشد.

نکته: روابط مفید زیر بین عملیات اصلی اجتماع، اشتراک و مکمل وجود دارد که به نام قوانین دمورگان شناخته می‌شوند:

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

اثبات: فرض کنید x نقطه‌ای از $\left(\bigcup_{i=1}^{n} E_i\right)^c$ باشد

$$x \in \left(\bigcup_{i=1}^{n} E_i\right)^c \Rightarrow x \notin \bigcup_{i=1}^{n} E_i \Rightarrow \underset{i=1,\dots,n}{\nexists i} \ni x \in E_i \Rightarrow \underset{i=1,\dots,n}{\forall i} \quad x \in E_i^c$$

$$\Rightarrow x \in \bigcap_{i=1}^{n} E_i^c$$

از طرف دیگر فرض کنیم x نقطه‌ای از $\bigcap_{i=1}^{n} E_i^c$ باشد، یعنی

$$x \in \bigcap_{i=1}^{n} E_i^c \Rightarrow \underset{i=1,\dots,n}{\forall i} \quad x \in E_i^c \Rightarrow \underset{i=1,\dots,n}{\nexists i} \ni x \in E_i$$

$$\Rightarrow x \notin \bigcup_{i=1}^{n} E_i \Rightarrow x \in \left(\bigcup_{i=1}^{n} E_i\right)^c$$

برای اثبات قانون دوم دمورگان از قانون اول دمورگان استفاده می‌کنیم.

$$\left(\bigcup_{i=1}^{n} E_i^c\right)^c = \bigcap_{i=1}^{n} (E_i^c)^c = \bigcap_{i=1}^{n} E_i$$

بنابراین $\left(\bigcup_{i=1}^{n} E_i^c\right)^c = \bigcap_{i=1}^{n} E_i$. حال با گرفتن متمم از طرفین تساوی داریم

$$\bigcup_{i=1}^{n} E_i^c = \left(\bigcap_{i=1}^{n} E_i\right)^c$$

اصول احتمال

یکی از راههای ممکن تعریف احتمال یک پیشامد بر اساس فراوانی نسبی است:

فرض کنید آزمایشی که فضای نمونه آن S است، تحت شرایط کاملاً یکسان تکرار شود. برای هر پیشامد E از فضای نمونه S، $n(E)$ را تعداد دفعاتی که پیشامد E در n مرتبه تکرار آزمایش اتفاق افتاده است در نظر می‌گیریم. آنگاه احتمال وقوع پیشامد E، $P(E)$، به صورت زیر تعریف می‌شود

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

یعنی $P(E)$ عبارت است از حد نسبت دفعاتی که E اتفاق می‌افتد و بنابراین حد فراوانی نسبی E است.

اگرچه تعریف فوق به طور قطع قابل قبول است، اما در ارتباط با پرسش زیر دارای نقطه ضعف جدی است.

" چگونه می‌دانیم که $\frac{n(E)}{n}$ به سمت یک مقدار ثابت حدی میل می‌کند که برای هر دنباله ای

از تکرارهای آزمایش یکسان است ؟ »

برای مثال فرض کنید آزمایشی که باید تکرار شود شامل پرتاب یک سکه باشد ، آنگاه چگونه می‌توانیم که نسبت شیرهای ظاهر شده در n پرتاب اولیه وقتی که n زیاد می‌شود به سمت یک مقدار میل می‌کند ؟ همچنین اگر این نسبت به سمت یک مقدار میل کند، چگونه مطمئن هستیم که اگر آزمایش برای بار دوم ، n مرتبه تکرار شود مجدداً نسبت حدی مثل را برای شیرها به دست می‌آوریم ؟

طرفداران تعریف فراوانی نسبی احتمال ، معمولاً پاسخ سؤالات فوق را با توجه به این نکته می‌دهند که همگرایی $\frac{n(E)}{n}$ به سمت یک مقدار ثابت یک فرض یا یک اصل است .

به هر حال فرض اینکه $\frac{n(E)}{n}$ لزوماً به یک مقدار ثابت حدی میل می‌کند فرض پیچیده‌ای است و ممکن است فقط امیدداشته باشیم که چنین حدی وجود دارد ، چون هیچ گونه شواهد قبلی که لزوم وجود حد را بیان کند در دست نیست .

در واقع منطقی‌تر است که مجموعه‌ای از اصول ساده‌تر را درباره احتمال بیان داشته و سپس ثابت کنیم که تعریف احتمال به صورت حد فراوانی نسبی وجود دارد . این کار ، روش اصولی و نوین نظریه احتمال نامیده شده است .

اصول اساسی احتمال

آزمایشی را در نظر بگیرید که فضای نمونه آن S است ، برای هر پیشامد E از فضای نمونه که فرض می‌کنیم $P(E)$ تعریف شده و سه اصل زیر را برآورده می‌کند :

اصل ۱ $\quad 0 \leq P(E) \leq 1$

اصل ۲ $\quad P(S) = 1$

اصل ۳ اگر E_1 ، E_2 ، . . . پیشامدهایی دو به دو ناسازگار باشند یعنی به ازای هر $i \neq j$

آنگاه $E_i \cap E_j = \emptyset$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

چند قانون احتمال

قضیه ۱: $P(\emptyset) = 0$

حل: در اصل سوم اصول موضوعه احتمال قرار دهیم $E_1 = S \qquad E_2 = E_3 = \dots = \emptyset$

بدیهی است که E_i ها از هم جدا هستند.

$$1 = P(S) = P\left(\bigcup_{i=1}^{\infty} E_i\right) \underset{\text{طبق اصل سوم}}{=} \sum_{i=1}^{\infty} P(E_i) = P(S) + \sum_{i=2}^{\infty} P(E_i) =$$

$$1 + \sum_{i=2}^{\infty} P(E_i) \quad \Rightarrow \sum_{i=2}^{\infty} P(E_i) = 0 \Rightarrow \sum_{i=2}^{\infty} P(\emptyset) = 0 \Rightarrow P(\emptyset) = 0$$

قضیه ۲: اگر F_1 ، F_2 ، ... ، F_n پیشامدهای دو به دو ناسازگار باشند، آنگاه

$$P(F_1 \cup F_2 \cup \dots \cup F_n) = \sum_{i=1}^{n} P(F_i)$$

اثبات: اگر در اصل سوم اصول موضوعه احتمال قرار دهیم $E_1 = F_1$ ، $E_2 = F_2$ ، ... ،

$E_n = F_n$ و نیز به ازای هر i بزرگتر از n قرار دهیم $E_i = \emptyset$ آنگاه

$$P(F_1 \cup F_2 \cup \dots \cup F_n) = P\left(\bigcup_{i=1}^{n} F_i\right) = P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

$$= \sum_{i=1}^{n} P(E_i) + \sum_{i=n+1}^{\infty} P(E_i) = \sum_{i=1}^{n} P(F_i) + \sum_{i=n+1}^{\infty} P(\emptyset) = \sum_{i=1}^{n} P(F_i)$$

پس برای اصل سوم اصول موضوعه احتمال برای یک دنباله متناهی به حجم n هم برقرار است.

مثال: اگر آزمایش شامل پرتاب یک سکه باشد و فرض کنیم آمدن شیر و خط هم شانس باشند، داریم

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}$$

از طرف دیگر اگر سکه اریب باشد و احساس کنیم که شانس آمدن شیر دو برابر خط است

آنگاه داریم $P(\{H\}) = \frac{2}{3} \qquad P(\{T\}) = \frac{1}{3}$

مثال: اگر تاسی را پرتاب و فرض کنیم که ۶ طرف آن شانس یکسان برای ظاهر شدن دارند،

داریم

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) =$$
$$P(\{6\}) = \frac{1}{6}$$

درحقیقت با توجه به اصل ۳ این نتیجه به دست می‌آید. احتمال اینکه نتیجه پرتاب یک تاس عدد زوج باشد برابر است با

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\})$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

نتیجه: اگر E و F دو پیشامد ناسازگار باشند آنگاه

$$P(E \cup F) = P(E) + P(F)$$

نتیجه: اگر E یک پیشامد و E^c متمم آن باشد آنگاه

$$P(E) + P(E^c) = 1 \Rightarrow P(E^c) = 1 - P(E)$$

اثبات: $E \cup E^c = S$ و $E \cap E^c = \emptyset$ بنابراین

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

مثال: اگر سکه‌ای را ۶ بار پرتاب کنیم آنگاه احتمال آوردن حداقل یک شیر را بیابید.

حل:

$$n(S) = 2^6 = 64$$

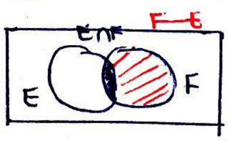
E: پیشامد مشاهده حداقل یک شیر E^c: پیشامد مشاهده هیچ شیر

$$E^c = \{TTTTTT\} \Rightarrow n(E^c) = 1$$

$$P(E^c) = \frac{n(E^c)}{n(S)} = \frac{1}{64} \Rightarrow P(E) = 1 - P(E^c) = \frac{63}{64}$$

نتیجه: اگر E و F دو پیشامد دلخواه باشند آنگاه

$$P(F-E) = P(F) - P(E \cap F)$$



$$(F-E) \cup (E\cap F) = F$$

$$(F-E) \cap (E\cap F) = \emptyset$$

بنابراین دو پیشامد ناسازگار $E\cap F$ و $F-E$ داریم و لذا

$$P(F) = P(F-E) + P(E\cap F)$$

نتیجه: اگر E و F دو پیشامد باشند که $E\subset F$ آنگاه

الف) $P(F-E) = P(F) - P(E)$ ب) $P(E) \leqslant P(F)$

برای قسمت الف ـ چون $E\subset F$ لذا $E\cap F = E$ و نتیجه بدیهی است.

برای قسمت ب ـ با توجه به آنکه تحت فرض $E\subset F$ داریم $E\cap F = E$ و لذا

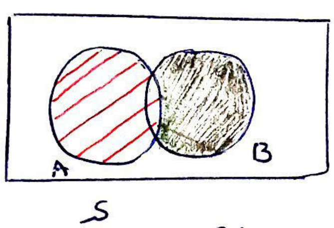
$$P(F-E) = P(F) - P(E)$$

از این رو طبق اصل اول اصول موضوعه احتمال چون $P(F-E) \geqslant 0$ لذا نتیجه حاصل می‌شود.

قضیه: اگر E و F دو پیشامد دلخواه باشند آنگاه

$$P(E\cup F) = P(E) + P(F) - P(E\cap F)$$

اثبات: همانطور که در نمودار مشاهده می‌شود دو پیشامد $F-E$ و E از هم جدا هستند و اجتماع آنها $E\cup F$ است لذا



$$P(E\cup F) = P(E) + P(F-E) = P(E) + P(F) - P(E\cap F)$$

نکته:

$$P(E\cup F\cup G) = P(E) + P(F) + P(G) - P(E\cap F) - P(E\cap G) - P(F\cap G) + P(E\cap F\cap G)$$

$$P(E\cup F\cup G) = P\{(E\cup F)\cup G\} = P(E\cup F) + P(G) - P((E\cup F)\cap G)$$

$$= P(E\cup F) + P(G) - P((E\cap G)\cup(F\cap G)) = P(E\cup F) + P(G) - P(E\cap G) - P(F\cap G) + P(E\cap F\cap G) = P(E) + P(F) + P(G) - P(E\cap F) - P(E\cap G) - P(F\cap G) + P(E\cap F\cap G)$$

گزاره ۵

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1} \cap E_{i_2}) + \dots$$

$$(-1)^{r+1} \sum_{i_1<i_2<\dots<i_r} P(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_r}) + \dots + (-1)^{n+1} P(E_1 \cap E_2 \cap \dots \cap E_n)$$

مجموع $\sum_{i_1<i_2<\dots<i_r} P(E_{i_1} \cap E_{i_2} \cdots E_{i_r})$ روی همه $\binom{n}{r}$ زیرمجموعه ممکن r تایی از مجموعه $\{1,2,\dots,n\}$ است.

بنابراین احتمال اجتماع n پیشامد برابر است با جمع احتمال تک تک پیشامدها منهای جمع احتمال اشتراک دو به دو پیشامدها، به علاوه جمع احتمال اشتراک سه تایی پیشامدها ...

فضای نمونه با نتایج هم‌شانس

آزمایشی را در نظر بگیرید که فضای نمونه آن مجموعه ای محدود مانند $S = \{1,2,\dots,N\}$ باشد، به گونه ای که

$$P(\{1\}) = P(\{2\}) = \dots = P(\{N\})$$

$$1 = P(S) = P(\{1\} \cup \{2\} \cup \dots \cup \{N\}) = P\{1\} + \dots + P\{N\}$$

$$= N\,P\{1\} \Rightarrow P\{1\} = \frac{1}{N}$$

بنابراین $P(\{i\}) = \frac{1}{N} \quad i = 1,\dots,N$

بنابراین برای هر پیشامد دلخواه مانند E، $E = \{1,\dots,n\}$ آنگاه

$$P(E) = P(\{1\} \cup \dots \cup \{n\}) = P\{1\} + \dots + P\{n\} = \frac{1}{N} + \dots + \frac{1}{N}$$

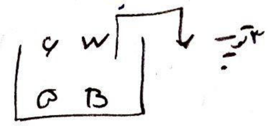
$$= \frac{n}{N} = \frac{\text{تعداد نقاط در } E}{\text{تعداد نقاط در } S}$$

مثال: اگر دو تاس را پرتاب کنیم، احتمال اینکه مجموع اعداد ظاهر شده برابر ۷ باشد چقدر است؟

E: پیشامد آنکه مجموع اعداد ظاهر شده برابر ۷ باشد

$$E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \qquad P(E) = \frac{6}{36} = \frac{1}{6}$$

مثال: اگر ۳ توپ را به تصادف از ظرفی که در آن ۶ توپ سفید و ۵ توپ سیاه است انتخاب کنیم. احتمال اینکه یکی از توپها سفید و دوتای دیگر سیاه باشند چقدر است؟



اگر ترتیب انتخاب توپها مهم باشد آنگاه فضای نمونه آزمایش دارای $11 \times 10 \times 9 = 990$ عضو خواهد بود. به علاوه به تعداد $6 \times 5 \times 4 = 120$ حالت امکان دارد که توپ اول سفید و دو توپ دیگر سیاه باشند و به تعداد $5 \times 6 \times 4 = 120$ حالت امکان دارد که توپ اول سیاه و توپ دوم سفید و توپ سوم سیاه باشد. همچنین به تعداد $5 \times 4 \times 6 = 120$ حالت امکان دارد که دو توپ اول سیاه و سومین توپ سفید باشد. با فرض انتخاب تصادفی هر یک از اعضای فضای نمونه دارای شانس برابر هستند.

$$\text{احتمال مورد نظر} = \frac{120+120+120}{990} = \frac{4}{11}$$

حال اگر ترتیب انتخاب توپها اهمیت نداشته باشد آنگاه

$$\text{احتمال مورد نظر} = \frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{4}{11}$$

مثال: یک شورای ۵ نفره را باید از یک گروه متشکل از ۶ مرد و ۹ زن انتخاب کرد. اگر انتخاب افراد تصادفی باشد احتمال اینکه شورا متشکل از ۳ مرد و ۲ زن باشد چقدر است؟

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001}$$

مثال: در ظرفی n توپ قرار دارد که یکی از آنها خاص است. اگر k توپ را از ظرف خارج کنیم به طوری که انتخاب هر توپ که در ظرف باقی می‌ماند شانس برابر با سایر توپها داشته باشد. مطلوبست احتمال اینکه توپ خاص از ظرف خارج شده باشد.

حل : روش اول :

$$P(\text{توپ خاص انتخاب شده باشد}) = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{\frac{(n-1)!}{(k-1)!}}{\frac{n!}{k!}} = \frac{k}{n}$$

روش دوم : A_i : پیشامد آنکه توپ خاص در i امین مرحله انتخاب شده باشد. $i=1,2,\dots,k$

$$P = P(\text{توپ خاص انتخاب شده باشد}) = P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i)$$

می دانیم تعداد نتایج هم شانس آزمایش برابر است با $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$

و تعداد نتایجی که توپ خاص در مرحله i انتخاب شده باشد برابر است با

$$(n-1)(n-2)\cdots(n-i+1)\ 1\ (n-i)\cdots(n-k+1) = \frac{(n-1)!}{(n-k)!}$$

$$P(A_i) = \frac{\frac{(n-1)!}{(n-k)!}}{\frac{n!}{(n-k)!}} = \frac{1}{n}$$

از این رو $P = \sum_{i=1}^{k} 1/n = k/n$

مثال : فرض کنید $n+m$ توپ را که n تای آنها قرمز و m تای آن آبی است در یک ردیف به ترتیب قرار دهیم به طوری که هر یک از $(n+m)!$ حالت ممکن قرار گرفتن توپها هم شانس باشند. اگر نتیجه مرتب کردن توپها را به صورت ترتیب رنگ توپها در نظر بگیریم نشان دهید که نتایج آزمایش بدین صورت نیز هم شانس می باشند.

حل : یکی از نتایج ممکن مرتب شدن توپها را از مجموع $(n+m)!$ حالت ممکن در نظر بگیریم. در این نتیجه ترتیب توپهای قرمز بین خودشان و همچنین ترتیب توپهای آبی بین خودشان تأثیری در تغییر نتیجه رنگ در این آزمایش ندارد. در نتیجه هر ترتیب از رنگ مربوط به $n!m!$ حالت مختلف مرتب شدن $n+m$ توپ می باشد که بنابراین احتمال هر ترتیب از رنگ

برابر $\frac{n!\,m!}{(n+m)!}$ است. برای مثال فرض کنید ۲ توپ قرمز r_1 و r_2 و دو توپ آبی b_1 و b_2 را در یک ردیف قرار می‌دهیم. تعداد حالتهای ممکن برابر با !۴ است که تعداد !۲!۲ از آنها نتیجه ترتیب رنگ می‌باشد. مثلاً نتایج زیر مربوط به حالتی است که اولین توپ قرمز باشد.

r_1, b_1, r_2, b_2 ، r_2, b_1, r_1, b_2 ، r_1, b_2, r_2, b_1 ، r_2, b_2, r_1, b_1

بنابراین احتمال هر یک از ترتیب رنگ برابر با $\frac{4}{24} = \frac{1}{6}$ است.

مثال: در ظرفی ۵۲ توپ از ۴ رنگ مختلف (قرمز، آبی، سبز و سفید) که هر رنگ با شماره‌های ۱ تا ۱۳ مشخص شده‌اند وجود دارد. از این ظرف ۵ توپ به تصادف و بدون جایگذاری انتخاب می‌کنیم. احتمال اینکه همه توپها از یک رنگ نبوده و اعداد روی آنها پشت سرهم باشد را به دست آورید.

حل: فرض کنید همه $\binom{52}{5}$ انتخاب توپها هم شانس باشند. تعداد حالاتی که اعداد ۱، ۲، ۳، ۴ و ۵ انتخاب می‌شوند برابر 4^5 است که ۴ حالت آن همه توپها از یک رنگ هستند. بنابراین 4^5-4 تعداد حالاتی است که ۵ توپ انتخاب شده از یک رنگ نبوده و اعداد آنها ۱، ۲، ۳، ۴ و ۵ باشد. به همین ترتیب برای اعداد ۲، ۳، ۴، ۵، ۶ نیز تعداد حالات مورد نظر برابر است با 4^5-4 و نهایتاً برای اعداد ۹، ۱۰، ۱۱، ۱۲، ۱۳ نیز تعداد حالات برابر 4^5-4 است که در مجموع تعداد کل حالات می‌شود برابر با $9(4^5-4)$ است و در نتیجه احتمال اینکه همه توپها از یک رنگ نبوده و اعداد روی آنها پشت سرهم باشد برابر است با

$$\frac{9(4^5-4)}{\binom{52}{5}} \approx 0/0036$$

مثال: در مثال قبل احتمال آن را به دست آورید که از ۵ توپ انتخاب شده ۳ توپ با شماره های یکسان و ۲ توپ دیگر نیز با شماره های یکسان باشند.

حل: فرض کنیم همه $\binom{52}{5}$ انتخاب توپ هم شانس باشند. می دانیم که به تعداد $\binom{4}{3}\binom{4}{2}$ حالت متمایز وجود دارد که ۳ توپ با شماره های یکسان (مثلاً شماره ۱۲) و ۲ توپ با شماره های دیگر یکسان (مثلاً شماره ۱۰) انتخاب شوند.

اما انتخاب ۳ توپ اول به ۱۳ طریق مختلف و انتخاب ۲ توپ بعدی به ۱۲ حالت مختلف امکان پذیر است و در نتیجه احتمال پیشامد مورد نظر برابر است با

$$\frac{13\times 12\times\binom{4}{3}\binom{4}{2}}{\binom{52}{5}} \approx 0/0014$$

مثال: در مثال قبل توپ ها را به تصادف بین ۴ نفر تقسیم می کنیم.

الف- احتمال اینکه یکی از افراد ۱۳ توپ هم رنگ را دریافت کند چقدر است؟

ب- احتمال اینکه هر یک از ۴ نفر توپ شماره ۵ را دریافت کند چقدر است؟

حل: الف- تعداد حالات ممکن برابر $\binom{52}{13,13,13,13}$ است و تعداد حالاتی که یک نفر ۱۳ توپ هم رنگ را دریافت کند برابر است با تعداد حالاتی که بقیه توپ ها بین سه نفر توزیع شود یعنی $\binom{39}{13,13,13}$. بنابراین احتمال مطلوب:

$$\frac{4\binom{39}{13,13,13}}{\binom{52}{13,13,13,13}} \approx 6/3\times 10^{-12}$$

ب-

برای محاسبه احتمال اینکه هر نفر توپ شماره ا را داشته باشد ابتدا این چهار توپ را کنار گذاشته و ۴۸ توپ باقی مانده را بین ۴ نفر توزیع می‌کنیم که به تعداد $\binom{48}{12,12,12,12}$ حالت امکان پذیر است. حال به تعداد !۴ حالت مختلف می‌توان ۴ توپ باقی مانده را بین بازیکن‌ها توزیع نمود و در نتیجه احتمال پیشامد مورد نظر برابر است با

$$\frac{4!\binom{48}{12,12,12,12}}{\binom{52}{13,13,13,13}} \simeq 0/105$$

مثال: اگر در اطاق n نفر باشند، احتمال اینکه هیچ دو نفری در یک روز از سال متولد نشده باشند را به دست آورید. n چقدر باشد تا این احتمال کمتر از $\frac{1}{2}$ باشد؟

چون هر فردی می‌تواند در هر یک از ۳۶۵ روز سال متولد شود، نتایج ممکن وقتی n نفر برابر با $(365)^n$ است. با فرض آنکه اسفند ۲۹ روز است و با فرض هم‌شانس بودن نتایج، احتمال مورد نظر برابر است با

$$\frac{365 \times 364 \times 363 \times \dots \times (365-(n-1))}{(365)^n}$$

تعجب آور است که برای $n \geq 23$ احتمال فوق کمتر از $\frac{1}{2}$ است. یعنی اگر ۲۳ نفر در اطاق باشند احتمال اینکه حداقل دو نفر در یک روز متولد شده باشند بیش از $\frac{1}{2}$ است. بسیاری این نتیجه را شگفت انگیز می‌دانند. زیرا به نظر می‌رسد که ۲۳ در مقایسه با ۳۶۵ عدد کوچکی است اگر چه برای هر دو نفر احتمال اینکه در یک روز به دنیا آمده باشند برابر با $\frac{1}{365} = \frac{365}{(365)^2}$ است.

ولی برای یک گروه ۲۳ نفری تعداد $253 = \binom{23}{2}$ زوج دو نفری وجود دارد که اگر بدین ترتیب به مسأله نگاه کنیم نتیجه به دست آمده دیگر جای تعجب ندارد. شاید این نتیجه که احتمال فوق به ۹۷/۰ برای ۵۰ نفر افزایش می‌یابد، حتی شگفت انگیزتر باشد. برای

۱۰۰ نفر احتمال اینکه حداقل دو نفر در یک روز متولد شده باشند بزرگتر از $\frac{3\times10^6}{3\times10^6+1}$ است.

درحقیقت

$$1-\prod_{i=0}^{99}(365-i)\Big/(365)^{100} > \frac{3\times10^6}{3\times10^6+1}$$

مثال: از طرف مثال کرمان ۵۲ توپ از ۴ رنگ مختلف قرمز، آبی، سفید و سبز است، به تصادف توپ را یک به یک خارج کرده و عدد آن را یادداشت می‌کنیم تا اولین توپ شماره ۱ ظاهر شود کدام یک از دو پیشامد: «آمدن توپ شماره ۱ قرمز بلافاصله بعد از اولین توپ شماره ۱»، و «آمدن توپ شماره ۲ آبی بلافاصله بعد از اولین توپ شماره ۱» محتمل‌تر است.

حل: برای محاسبه احتمال اینکه توپ شماره ۱ قرمز بلافاصله بعد از مشاهده اولین توپ شماره ۱ ظاهر شود نیاز به محاسبه تعداد حالتهای از !۵۲ حالت ممکن که پیشامد مورد نظر حاصل شود داریم. بدین منظور اگر !۵۲ حالت را بدین صورت در نظر بگیریم که ابتدا تعداد ترتیب های ۵۱ توپ بدون توپ شماره ۱ قرمز را محاسبه نموده و سپس توپ شماره ۱ قرمز را در داخل ترتیب ها قرار دهیم. بدیهی است که برای هر یک از !۵۱ حالت، فقط یک حالت وجود دارد که توپ شماره ۱ قرمز بلافاصله بعد از اولین توپ شماره ۱ ظاهر شود.

$$P\left(\text{توپ شماره ۱ قرمز بعد از اولین توپ شماره ۱}\right) = \frac{51!}{52!} = \frac{1}{52}$$

درواقع این استدلال برای توپ شماره ۲ آبی نیز صادق است؛ به عبارتی برای هر شماره ای از ۴ رنگ این احتمال برابر با $\frac{1}{52}$ می باشد. به بیان دیگر هر یک از ۵۲ توپ دارای شانس برابر هستند که بعد از اولین توپ شماره ۱ ظاهر شوند.

بسیاری این نتیجه را تعجب آور می دانند درواقع انتظار عموم این است که آمدن توپ آبی شماره ۲ محتمل‌تر است. زیرا اولین توپ شماره ۱ می‌تواند توپ شماره ۱ قرمز باشد. پاسخ این تفکر چنین است که توپ آبی شماره ۲ هم می‌تواند قبل از اولین توپ شماره ۱ ظاهر شود در نتیجه شانس اینکه بعد از اولین توپ شماره ۱ ظاهر شود را کاهش دهد. از طرف

چون شانس ۱ به ۴ برای ظاهر شدن توپ شماره ۱ قرمز به عنوان اولین توپ شماره ۱ و شانس ۱ به ۵ برای ظاهر شدن توپ شماره ۲ آبی قبل از اولین توپ شماره ۱ [زیرا از هر مجموعه ۵ توپی که توپ شماره ۲ آبی و ۴ توپ شماره ۱ دیگر در آن باشد یک حالت (با شانسهای برابر) وجود دارد که توپ شماره ۲ آبی پیش از ۴ توپ شماره ۱ ظاهر شود] وجود دارد. بنابراین بازای مقدار ایجاد می‌شود که ظاهر شدن توپ شماره ۲ آبی محتمل‌تر است و این در حالی است که نتیجه حل مسأله بیان می‌دارد که هر دو پیشامد هم شانس هستند. □

مثال ۵-۱۱ یک تیم فوتبال دارای ۲۰ بازیکن حمله و ۲۰ بازیکن دفاع است. بازیکن‌ها باید در گروه‌های دوتایی برای تعیین هم اتاق خود تقسیم شوند. اگر زوج‌ها به طور تصادفی انتخاب شوند، احتمال اینکه در هیچ زوجی، بازیکن دفاع - حمله نباشد چقدر است؟ احتمال اینکه $2i$ زوج بازیکن دفاع - حمله تشکیل شوند چقدر است؟ $(i=1,2,\dots,10)$

حل: تعداد گروه‌های ۲ نفره، از تیم ۴۰ نفر برابر است با

$$\binom{40}{2,2,\dots,2} = \frac{40!}{(2!)^{20}}$$

(پس به تعداد $\frac{40!}{2^{20}}$ طریق مختلف می‌توان بازیکن‌ها را به زوج اول، زوج دوم ... تقسیم کرد) بنابراین به تعداد $\frac{40!}{20!\,2^{20}}$ طریق مختلف می‌توان آنها را به زوج‌های مرتب نشده تقسیم کرد. به علاوه چون تقسیمی که نتیجه آن هیچ زوج بازیکن حمله - دفاع نباشد از تشکیل زوج‌های دفاع و زوج‌های حمله بین خودشان به وجود می‌آید و تعداد آن برابر $\left[\frac{20!}{2^{10}10!}\right]^2$ است، پس احتمال اینکه هیچ زوج بازیکن دفاع - حمله تشکیل نشود برابر است با

$$P_0 = \frac{\left[\frac{20!}{2^{10}10!}\right]^2}{\frac{40!}{2^{20}\,20!}} = \frac{(20!)^3}{(10!)^2\,40!}$$

برای محاسبه P_{2i}، احتمال اینکه $2i$ زوج بازیکن دفاع - حمله تشکیل شود ابتدا توجه می‌کنیم که به تعداد $\binom{20}{2i}^2$ طریق مختلف می‌توان $2i$ بازیکن دفاع و

$2i$ بازیکن حمله را انتخاب نمودها ین یا 4 بازیکن می توانند $(2i)!$ مختلف تشکیل زوج بازیکن دفاع - حمله بدهند ، زیرا اولین بازیکن دفاع می تواند با هر یک از $2i$ بازیکن حمله ، دومین بازیکن دفاع با هر یک از $(2i-1)$ بازیکن باقی مانده حمله و... تشکیل زوج دهند . باقی مانده $20-2i$ بازیکن (دفاع و حمله) باید بین خودشان تشکیل زوج دهند و در نتیجه تعداد حالتهای ممکن تشکیل $2i$ زوج بازیکن دفاع - حمله برابر است با

$$\binom{20}{2i}^2 (2i)! \left[\frac{(20-2i)!}{2^{10-i}(10-i)!}\right]^2$$

$$\Rightarrow P_{2i} = \frac{\binom{20}{2i}^2 (2i)! \left[\frac{20-2i}{2^{10-i}(10-i)!}\right]^2}{\frac{40!}{2^{20}\,20!}} \qquad i=0,1,2,\ldots,10$$

با توجه به تقریب استرلینگ

$$n! \simeq n^{n+1/2} e^{-n} \sqrt{2\pi}$$

داریم

$$P_0 \approx 1{,}3403 \times 10^{-6} \qquad P_{10} \approx 0{,}345861$$

$$P_{20} \approx 7{,}6068 \times 10^{-6}$$

مثال : در یک باشگاه ورزشی ، 36 نفر بازی تنیس ، 28 نفر اسکواش ، 18 نفر بدمینتون ، 22 نفر تنیس و اسکواش ، 12 نفر تنیس و بدمینتون ، 9 نفر اسکواش و بدمینتون و 4 نفر هر سه ورزش را انجام می دهند. چه تعداد از اعضای این باشگاه حداقل یک بازی را انجام می دهند .

حل : اگر N نشان دهنده تعداد اعضای باشگاه باشد ، با فرض انتخاب تصادفی یک بازیکن ، هر زیرمجموعه C از اعضای باشگاه دارای احتمال

$$P(C) = \frac{\text{تعداد افرادی که در زیرمجموعه } C \text{ هستند}}{N}$$

اگر T نشان دهنده مجموعه اعضایی که بازی تنیس می کنند ، S کسانی که بازی اسکواش و B

عده ای که بازی مورد علاقه شان را انجام می دهند باشد آنگاه

$$P(T\cup S\cup B) = P(T)+P(S)+P(B)-P(T\cap S)-P(T\cap B)$$

$$-P(S\cap B)+P(T\cap S\cap B) = \frac{36+28+18-22-12-9+4}{N} = \frac{43}{N}$$

بنابراین تعداد ۴۳ نفر از بازیکنان حداقل یک ورزش را انجام می دهند.

مسأله انطباق فرض کنید هر یک از N مردی که در یک میهمانی شرکت دارند کلاه خود را در وسط اتاق پرتاب کنند. آنگاه ابتدا کلاه ها را مخلوط نموده و سپس هر مرد به تصادف کلاهی را انتخاب کند.

الف - احتمال اینکه هیچ مردی کلاه خود را انتخاب نکند چقدر است؟

ب - احتمال اینکه دقیقاً K مرد کلاه خود را انتخاب کنند چقدر است؟

حل: الف) ابتدا احتمال متمم پیشامد را که حداقل یک نفر کلاه خود را انتخاب کند محاسبه می کنیم. فرض کنید

E_i : پیشامد آنکه i امین مرد کلاه خود را انتخاب کند $i=1,\dots,N$

$$P\left(\bigcup_{i=1}^{N}E_i\right) = \sum_{i=1}^{N}P(E_i)-\sum_{i_1<i_2}P(E_{i_1}\cap E_{i_2})+\dots+(-1)^{n+1}\sum_{i_1<i_2<\dots<i_n}P(E_{i_1}\cap E_{i_2}\cap\dots\cap E_{i_n})$$

$$+\dots+(-1)^{N+1}P(E_1\cap E_2\cap\dots\cap E_N)$$

اگر نتیجه این آزمایش به صورت برداری از N عدد باشد که i امین عضو آن، عدد کلاه انتخاب شده توسط مرد i ام باشد آنگاه تعداد نتایج ممکن برابر $N!$ است. نتیجه $(1,2,3,\dots,N)$ نشان دهنده پیشامدی است که همه مردها کلاه خودشان را انتخاب کنند. همچنین اگر $E_{i_1}E_{i_2}\dots E_{i_n}$ نشان دهنده پیشامدی باشد که هر یک از n مرد $i_1, i_2, \dots, i_n$ کلاه خودش را انتخاب کند تعداد حالت های ممکن برابر است با

$$(N-n)(N-n-1)\cdots 3\times 2\times 1 = (N-n)!$$

زیرا برای $N-n$ مرد باقی مانده، اولین نفر می تواند هر یک از $N-n$ کلاه را انتخاب

کنند دومین نفر هر یک از $N-n-1$ کلاه باقی مانده و .... بنابراین با فرض اینکه همه

این N نتیجه خاص هم شانس باشند داریم

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}) = \frac{(N-n)!}{N!}$$

از طرفی تعداد $\binom{N}{n}$ جمله در $\sum_{i_1<i_2<\cdots<i_n} P(E_{i_1} \cap \cdots E_{i_n})$ وجود دارد، پس

$$\sum_{i_1<i_2<\cdots<i_n} P(E_{i_1} \cap \cdots \cap E_{i_n}) = \frac{N!\,(N-n)!}{(N-n)!\,n!\,N!} = \frac{1}{n!}$$

$$P\left(\bigcup_{i=1}^{N} E_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{N+1}}{N!}$$

و احتمال اینکه هیچ مردی کلاه خود را انتخاب نکند برابر است با

$$1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^N}{N!}$$

که برای N بزرگ تقریباً برابر $e^{-1} \approx ۰/۳۶۷۸۸$ است. به عبارت دیگر برای N بزرگ احتمال اینکه هیچ مردی کلاه خود را انتخاب نکند تقریباً ۰/۳۷ است.

ب − برای محاسبه احتمال اینکه دقیقاً k نفر از N مرد کلاه خود را انتخاب کنند، ابتدا k نفر را انتخاب نموده، سپس احتمال آن را محاسبه می‌کنیم که این k نفر کلاه خود را انتخاب کنند و سپس حاصل را در احتمال آنکه $N-k$ مرد باقیمانده کلاه خود را انتخاب نکنند ضرب می‌کنیم.

$$\binom{N}{k}\left(\frac{1}{N} \times \frac{1}{N-1} \times \cdots \times \frac{1}{N-(k-1)}\right)\left(\frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-k}}{(N-k)!}\right)$$

$$= \binom{N}{k}\frac{(N-k)!}{N!}\left(\frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-k}}{(N-k)!}\right)$$

$$= \frac{N! \times (N-k)!}{k!\,(N-k)!\,N!}\left(\frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-k}}{(N-k)!}\right)$$

$$= \frac{\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^{N-k}}{(N-k)!}}{k!}$$

که وقتی N بزرگ باشد تقریباً برابر $\frac{e^{-1}}{k!}$ است. مقدار $\frac{e^{-1}}{k!}$ از جنبه نظری دارای اهمیت است، زیرا مقادیری در توزیع پواسون با پارامتر یک است.

مثال: اگر ۱۰ زوج متأهل دور یک میز گرد غذاخوری نشسته باشند احتمال اینکه هیچ زن در کنار شوهرش نباشد را به دست آورید.

E_i : نشان دهنده پیشامد آنکه iامین زوج کنار هم نشسته باشند $\quad i=1,2,\dots,10$

احتمال مطلوب عبارت است از $\quad 1-P(\bigcup_{i=1}^{10} E_i)$

$$P(\bigcup_{i=1}^{10} E_i) = \sum_{i=1}^{10} P(E_i) - \dots + (-1)^{n+1} \sum_{i_1<i_2<\dots<i_n} P(E_{i_1}\cap E_{i_2}\cap\dots\cap E_{i_n})$$

$$+ \dots - P(E_1\cap E_2\cap\dots\cap E_{10})$$

ابتدا توجه کنید که ۲۰ نفر می‌توانند به $19!$ حالت مختلف دور یک میز گرد بنشینند.

$$P(E_{i_1}\cap E_{i_2}\cap\dots\cap E_{i_n}) = \frac{2^n(19-n)!}{19!}$$

بنابراین احتمال اینکه حداقل یک زوج کنار هم قرار گیرند برابر است با

$$\binom{10}{1}2^1\frac{18!}{19!} - \binom{10}{2}2^2\frac{17!}{19!} + \binom{10}{3}2^3\frac{16!}{19!} - \dots - \binom{10}{10}2^{10}\frac{9!}{19!}$$

≈ 0.6605

و احتمال موردنظر تقریباً برابر 0.3395 است.

مثال: دنباله‌ها یک تیم ورزشی را در نظر بگیرید که اخیراً فصل بازیهای آن تمام شده است و نتیجه $n+m$ بازی، n برد و m باخت بوده است. با بررسی دنباله بردها و باختها امیدواریم بتوانیم پاسخ این سؤال که آیا در بازیهای دوره شانس برد بیشتر است را تعیین کنیم. یک راه پاسخ دادن به سؤال این است که تعداد دنباله‌های برد را شمرده و ببینیم که شانس مورد نظر چقدر است؛ با توجه به اینکه $\frac{(n+m)!}{n!m!}$ ترتیب مختلف n برد و m باخت هم شانس باشند. منظور از یک دنباله برد این است که بردهای متوالی را در

نظر بگیریم. مثلاً اگر $m=6$ و $n=10$ و دنباله نتایج به صورت

WWLLWWWLWLLLWWWW

باشد، ۴ دنباله برد داریم که اندازه اولین دنباله ۲، دومین دنباله ۳، سومین دنباله ۱ و چهارمین دنباله ۴ است.

حال فرض کنید که یک تیم n برد و m باخت داشته باشد و همه $\binom{n+m}{n} = \frac{(n+m)!}{n!m!}$ ترتیب هم شانس باشند. برای محاسبه احتمال اینکه دقیقاً r دنباله از بردها وجود دارد ابتدا بردار اعداد صحیح و مثبت $x_1, x_2, \dots, x_r$ با شرط $x_1+x_2+\dots+x_r=n$ را در نظر گرفته و تعیین می‌کنیم که نتایج به چند حالت می‌توانند در r دنباله ظاهر شوند، به طوری که دنباله iام دارای اندازه x_i $(i=1,\dots,r)$ باشد. برای چنین نتیجه‌ای اگر y_1 نشان دهنده تعداد باخت‌ها قبل از اولین دنباله بردها، y_2 تعداد باخت‌ها بین ۲ دنباله اول بردها و ... y_{r+1} تعداد باخت‌ها بعد از آخرین دنباله بردها باشد، داریم

$$y_1+y_2+\dots+y_{r+1}=m \qquad y_1\geq 0 \qquad y_{r+1}\geq 0 \qquad y_i>0 \quad i=2,\dots,r$$

مطلب مزبور را می‌توان به صورت زیر نمایش داد:

$$\underbrace{LL\dots L}_{y_1}\ \underbrace{WW\dots W}_{x_1}\ \underbrace{L\dots L}_{y_2}\ \underbrace{WW\dots W}_{x_2}\dots\underbrace{W\dots W}_{x_r}\ \underbrace{L\dots L}_{y_{r+1}}$$

بنابراین تعداد نتایجی که منجر به r دنباله برد با اندازه x_i $(i=1,\dots,r)$ می‌شود برابر است با تعداد اعداد صحیح $y_1, y_2, \dots, y_{r+1}$ که در رابطه فوق صدق کنند و یا به عبارت دیگر تعداد اعداد صحیح مثبت

$$\bar{y}_1 = y_1+1 \qquad \bar{y}_i = y_i \quad i=2,\dots,r$$

$$\bar{y}_{r+1} = y_{r+1}+1$$

که

$$\bar{y}_1+\bar{y}_2+\dots+\bar{y}_{r+1} = m+2$$

در نتیجه تعداد چنین نتایجی برابر $\binom{m+1}{r}$ است. بنابراین تعداد کل نتایج که r دنباله از بردها را نتیجه می‌دهد برابر $\binom{m+1}{r}$ ضرب در تعداد اعداد صحیح مثبت حل معادله $x_1+x_2+\cdots+x_r=n$ است.

$$P(\{\text{r دنباله از بردها}\}) = \frac{\binom{m+1}{r}\binom{n-1}{r-1}}{\binom{m+n}{n}} \qquad r \geq 1$$

به عنوان مثال وقتی $m=6$ و $n=8$، احتمال ۷ دنباله برابر است با:

$$\frac{\binom{7}{7}\binom{7}{6}}{\binom{14}{8}} = \frac{1}{429}$$

در صورتی که همه $\binom{14}{8}$ نتیجه ممکن هم شانس باشند. بنابراین اگر نتیجه به صورت WLWLWLWLWLWWLWLW باشد ممکن است مشکوک شویم که احتمال بردن تیم در طول زمان تغییر می‌کند (به خصوص احتمال بردن تیم وقتی که بازی قبل را باخته باشد به نظر زیاد است وقتی که بازی قبل را برده باشد کاملاً اندک است) از طرف دیگر اگر نتیجه به صورت WWWWWWWWLLLLLL باشد، فقط یک دنباله وجود دارد

$$P(\{\text{یک دنباله}\}) = \frac{\binom{7}{1}\binom{7}{0}}{\binom{14}{8}} = \frac{1}{429}$$

که مجدداً به نظر می‌رسد احتمال برد تیم در ۱۴ بازی خود تغییر نکرده است.

احتمال به عنوان یک تابع مجموعه‌ای پیوسته

دنباله پیشامدهای $\{E_n, n \geq 1\}$ را یک دنباله صعودی نامیم هرگاه

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

و یک دنباله نزولی است هرگاه

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

اگر $\{E_n, n\geq 1\}$ یک دنباله صعودی از پیشامدها باشد، آنگاه پیشامد جدیدی را که با $\lim\limits_{n\to\infty} E_n$ نمایش می‌دهیم به صورت زیر تعریف می‌کنیم

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

به طور مشابه، اگر $\{E_n, n\geq 1\}$ یک دنباله نزولی از پیشامدها باشد $\lim\limits_{n\to\infty} E_n$ را به شکل زیر تعریف می‌کنیم

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

گزاره: اگر $\{E_n, n\geq 1\}$ یک دنباله صعودی یا نزولی از پیشامدها باشد آنگاه

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

اثبات: ابتدا فرض می‌کنیم که $\{E_n, n\geq 1\}$ یک دنباله صعودی باشد و پیشامد F_n $\{n\geq 1\}$ را به صورت زیر تعریف می‌کنیم

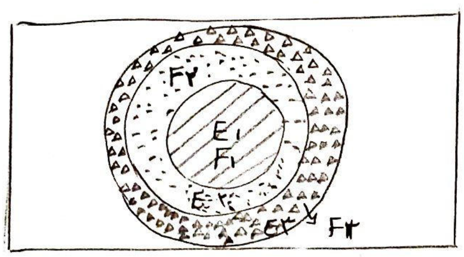
$$F_1 = E_1$$

$$F_2 = E_2 - E_1 = E_2 \cap E_1^c$$

$$F_3 = E_3 - E_2 = E_3 \cap E_2^c = E_3 \cap \left(\bigcup_{i=1}^{2} E_i\right)^c$$

$$\vdots$$

$$F_n = E_n \cap \left(\bigcup_{i=1}^{n-1} E_i\right)^c = E_n \cap E_{n-1}^c = E_n - E_{n-1} \qquad n>1$$



F_nها دو به دو ناسازگارند به طوری که برای هر $n\geq 1$

$$\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i$$

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i$$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = P\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P(F_i)$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} P(F_i) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} F_i\right) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} E_i\right) = \lim_{n\to\infty} P(E_n)$$

که نتیجه برای $\{E_n, n\geq 1\}$ وقتی که صعودی است ثابت می‌شود. اگر $\{E_n, n\geq 1\}$ دنباله‌ای

نزولی از پیشامدها باشد، آنگاه $\{E_n^c ,\, n\geq 1\}$ دنباله‌ای صعودی است و بنابراین با استفاده از رابطه قبل داریم

$$P\left(\bigcup_{i=1}^{\infty} E_i^c\right) = \lim_{n\to\infty} P(E_n^c)$$

$$P\left(\bigcup_{i=1}^{\infty} E_i^c\right) = P\left(\left(\bigcap_{i=1}^{\infty} E_i\right)^c\right) = 1 - P\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{n\to\infty}\left[1 - P(E_n)\right]$$

$$= 1 - \lim_{n\to\infty} P(E_n) \Rightarrow \lim_{n\to\infty} P(E_n) = 1 - P\left(\bigcup_{i=1}^{\infty} E_i^c\right) = P\left(\bigcap_{i=1}^{\infty} E_i\right)$$

مثال: ثابت کنید

$$\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$$

داریم

$$k\times\binom{n}{k} = k\times\frac{n!}{k!(n-k)!} = n\frac{(n-1)!}{(k-1)!(n-k)!} = n\binom{n-1}{k-1}$$

بنابراین

$$\sum_{k=1}^{n} k\binom{n}{k} = \sum_{k=1}^{n} n\binom{n-1}{k-1} = n\sum_{k=0}^{n-1}\binom{n-1}{k}1^k\,1^{n-1-k}$$

$$= n(1+1)^{n-1} = n2^{n-1}$$

تست: کدام تساوی درست نیست؟

۱) $\sum_{k=1}^{n}\binom{n}{k} = 2^n - 1$ ۲) $\binom{n}{r} = \binom{n}{n-r}$

۳) $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$ ۴) $1\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \dots + n\binom{n}{n} = n2^n$ ✓

تست: در بسط $(x_1+x_2+\dots+x_p)^n$ چند جمله وجود دارد؟

۱) $\binom{n+p-1}{p}$ ۲) $\binom{n+p}{p}$ ۳) $\binom{n+p}{n}$ ۴) $\binom{n+p-1}{n}$

$\binom{n+p-1}{n}$ تعداد جوابهای صحیح نامنفی معادله $n_1+n_2+\dots+n_p = n$

تست: به چند روش می‌توان ۳۰ مهره نامتمایز را در ۱۵ ظرف متمایز قرار داد به طوری که ظرف شماره ۷ درست ۱۴ مهره داشته باشد؟

۱) $\binom{29}{14}$ ۲) $\binom{30}{14}$ ۳) $\binom{29}{13}$ ✓ ۴) $\binom{30}{13}$

$x_1+x_2+\dots+x_{14} = 16$ $x_i \geq 0$ $\binom{16+14-1}{14-1} = \binom{29}{13}$

تمرین: اگر $P(E)=0.9$ و $P(F)=0.8$ باشد، نشان دهید که $P(E\cap F)\geq 0.7$ است. به طور کلی

نامساوی زیر که به نامساوی بونفرونی مشهور است را اثبات کنید.

$$P(E\cap F)\geq P(E)+P(F)-1$$

$$P(E\cup F)\leq 1 \Rightarrow P(E)+P(F)-P(E\cap F)\leq 1 \Rightarrow P(E\cap F)\geq P(E)+P(F)-1$$

لذا در مثال فوق $P(E\cap F)\geq 0.9+0.8-1=0.7$

تمرین: با استفاده از استقراء ریاضی، تعمیم نامساوی بونفرونی برای n پیشامد را اثبات کنید

$$P(E_1\cap E_2\cap\dots\cap E_n)\geq P(E_1)+\dots+P(E_n)-(n-1)$$

به ازای $n=2$ در تمرین قبل ثابت شد. فرض می‌کنیم نامعادله به ازای $n-1$ برقرار است و به ازای n آن را ثابت می‌کنیم

$$P(E_1\cap E_2\cap\dots\cap E_n)=P[(E_1\cap E_2\cap\dots\cap E_{n-1})\cap E_n]$$

$$\geq P(E_1\cap E_2\cap\dots\cap E_{n-1})+P(E_n)-1\geq P(E_1)+\dots+P(E_{n-1})-((n-1)-1)$$

$$+P(E_n)-1=P(E_1)+P(E_2)+\dots+P(E_n)-(n-1)$$

تمرین: نامساوی بول را اثبات کنید

$$P(\bigcup_{i=1}^{\infty}E_i)\leq\sum_{i=1}^{\infty}P(E_i)$$

اثبات با استفاده از روش استقراء.

ابتدا نشان می‌دهیم که این نامساوی برای اجتماع‌های متناهی و محدود برقرار است، به عبارت دیگر ادعای زیر را اثبات می‌کنیم

$$P(\bigcup_{k=1}^{n}E_n)\leq\sum_{k=1}^{n}P(E_k)$$

با توجه به آنکه $P(E_1)\leq P(E_1)$ لذا نامساوی به ازای $n=1$ برقرار است. حال فرض می‌کنیم نامساوی به ازای $n=m$ برقرار است، به عبارت دیگر فرض می‌کنیم

$$P(\bigcup_{k=1}^{m}E_k)\leq\sum_{k=1}^{m}P(E_k)$$

برقرار است. باید نشان می‌دهیم عبارت به ازای $n=m+1$ برقرار است، به بیان دیگر نشان می‌دهیم

$$P(\bigcup_{k=1}^{m+1}E_k)\leq\sum_{k=1}^{m+1}P(E_k)$$

حال

$$P\left(\bigcup_{k=1}^{m+1} E_k\right) = P\left(\bigcup_{k=1}^{m} E_k \cup E_{m+1}\right) = P\left(\bigcup_{k=1}^{m} E_k\right) + P(E_{m+1}) - P\left(\left(\bigcup_{k=1}^{m} E_k\right) \cap E_{m+1}\right)$$

$$\leq P\left(\bigcup_{k=1}^{m} E_k\right) + P(E_{m+1}) - 0 = P\left(\bigcup_{k=1}^{m} E_k\right) + P(E_{m+1}) \leq \sum_{k=1}^{m} P(E_k) + P(E_{m+1}) = \sum_{k=1}^{m+1} P(E_k)$$

بنابراین توانستیم با استفاده از استقراء نشان دهیم که نامساوی زیر

$$P\left(\bigcup_{k=1}^{n} E_k\right) \leq \sum_{k=1}^{n} P(E_k)$$

برای تمامی اعداد صحیح مثبت n برقرار است.

حال چون $\left\{\bigcup_{k=1}^{n} E_k\right\}$ یک دنباله صعودی از پیشامدها را تشکیل می‌دهند لذا

$$P\left(\bigcup_{k=1}^{\infty} E_k\right) = P\left(\lim_{n\to\infty} \bigcup_{k=1}^{n} E_k\right) = \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} E_k\right) \leq \lim_{n\to\infty} \sum_{k=1}^{n} P(E_k)$$

$$= \sum_{k=1}^{\infty} P(E_k)$$

Boole's inequality

روش دوم برای اثبات نامساوی بول: ما ابتدا یک گردایه مجزا از دنباله‌های $F_1, F_2, \dots$ با این خاصیت می‌سازیم که $\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} F_i$. تعریف می‌کنیم

$$F_1 = E_1 \qquad F_2 = E_2 - E_1 \qquad F_3 = E_3 - (E_1 \cup E_2), \dots$$

یا به طور کلی قرار می‌دهیم $F_i = E_i - \left(\bigcup_{j=1}^{i-1} E_j\right) \quad i = 2, 3, \dots$

به سادگی می‌توان بررسی نمود که $\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i$ است و بنابراین

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = P\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P(F_i)$$

زیرا F_i ها از هم مجزا هستند: در واقع

$$F_i \cap F_k = \left(E_i \cap \left(\bigcup_{j=1}^{i-1} E_j\right)^c\right) \cap \left(E_k \cap \left(\bigcup_{j=1}^{k-1} E_j\right)^c\right)$$

$$= \left[E_i \cap \left(\bigcap_{j=1}^{i-1} E_j^c\right)\right] \cap \left[E_k \cap \left(\bigcap_{j=1}^{k-1} E_j^c\right)\right]$$

طبق قانون دمورگان

حال اگر $k < i$ ، اولین اشتراک فوق در مجموعه E_k^c قرار داده می‌شود که دارای اشتراک تهی با E_k است. اگر $i > k$ ، بحث مشابه است. علاوه بر این با توجه به آنکه $F_i \subset E_i$ بنابراین $P(F_i) \leqslant P(E_i)$ و لذا داریم

$$\sum_{i=1}^{\infty} P(F_i) \leqslant \sum_{i=1}^{\infty} P(E_i)$$

در نتیجه مطلوب حاصل می‌شود.

مثال: احتمال و تیپ پارادوکس فرض کنید ظرفی با گنجایش نامحدود داریم و مجموعه‌ای نامحدود از توپ‌ها را که با شماره‌های ۱، ۲، ... شماره گذاری کرده‌ایم در اختیار داریم. آزمایش را به این صورت انجام می‌دهیم که در ۱ دقیقه به ساعت ۱۲ نیمه شب، توپ‌های شماره ۱ تا ۱۰ را در ظرف قرار داده و توپ شماره ۱۰ را خارج می‌کنیم. (فرض کنید خارج کردن توپ زمان نمی‌برد) در $\frac{1}{2}$ دقیقه به ساعت ۱۲ نیمه شب توپ‌های شماره ۱۱ تا ۲۰ را در ظرف قرار داده و توپ شماره ۲۰ را خارج می‌کنیم. در $\frac{1}{4}$ دقیقه به ساعت ۱۲ نیمه شب توپ‌های شماره‌های ۲۱ تا ۳۰ را در ظرف قرار داده و توپ شماره ۳۰ را خارج می‌کنیم. در $\frac{1}{8}$ دقیقه به ساعت ۱۲ نیمه شب و ...

سوال جالب این است که در ساعت ۱۲ چند توپ در ظرف است؟ پاسخ این سوال به طور واضح این است که بی‌نهایت توپ در ساعت ۱۲ در ظرف قرار دارد. زیرا هر توپی که شماره آن به صورت $10n$ $(n \geqslant 1)$ نباشد در داخل ظرف قرار گرفته و تا ساعت ۱۲ از آن خارج نمی‌شود. بنابراین اگر آن طور که شرح داده شد آزمایش انجام گیرد، مسئله حل شده است. اما اگر آزمایش را بدین ترتیب تغییر دهیم که ۱ دقیقه به ساعت ۱۲، توپ‌های شماره ۱ تا ۱۰ را در ظرف قرار داده و توپ شماره ۱ را خارج می‌کنیم، در $\frac{1}{2}$ دقیقه به ۱۲، توپ‌های شماره ۱۱ تا ۲۰ را قرار داده و توپ شماره ۲ را خارج می‌کنیم. در $\frac{1}{4}$ دقیقه به ساعت ۱۲: توپ‌های ۲۱ تا ۳۰ را قرار داده و

و توپ شماره ۳ را خارج می‌کنیم و در $\frac{1}{8}$ دقیقه به ساعت ۱۲، توپهای ۳۱ تا ۴۰ را در ظرف قرار داده، توپ شماره ۴ را خارج می‌کنیم و ... . برای این آزمایش جدید در ساعت ۱۲ چه تعداد توپ در ظرف وجود دارد؟

با کمال تعجب پاسخ این است که در ساعت ۱۲ نیمه‌شب ظرف خالی است. زیرا هر توپی را که در نظر بگیریم، مثلاً توپ شماره n در زمانی پیش از ۱۲ (در $\left(\frac{1}{2}\right)^{n-1}$ دقیقه به ۱۲) از ظرف خارج شده است. بنابراین برای هر n، توپ شماره n در ساعت ۱۲ در ظرف نیست. یعنی ظرف در این ساعت خالی است. بنابراین چگونگی خارج کردن توپ از ظرف موجب اختلاف در پاسخ به سؤال می‌شود. حال فرض کنید هر توپی را که می‌خواهیم از ظرف خارج کنیم، به طور تصادفی از توپهای موجود در ظرف انتخاب کنیم. در این حالت چند توپ در ساعت ۱۲ در ظرف باقی می‌ماند؟

حل: نشان خواهیم داد که با احتمال ۱، ظرف در ساعت ۱۲ نیمه‌شب خالی است. ابتدا توپ شماره ۱ را در نظر بگیریم. اگر E_n نشان دهنده پیشامدی باشد که توپ شماره ۱ بعد از خارج کردن n توپ هنوز در ظرف باقی مانده باشد داریم

$$P(E_n) = \frac{9 \times 18 \times 27 \times \cdots \times (9n)}{10 \times 19 \times 28 \times \cdots \times (9n+1)}$$

(برای درک این نتیجه توجه کنید که اگر بخواهیم توپ شماره ۱ بعد از n انتخاب در ظرف باقی بماند، باید در انتخاب اولین توپ هر یک از ۹ توپ دیگر انتخاب شود، در انتخاب دومین توپ هر یک از ۱۸ توپ دیگر موجود در ظرف انتخاب شوند و ...)

حال پیشامد اینکه توپ شماره ۱ در ساعت ۱۲ نیمه‌شب در ظرف باشد به صورت $\bigcap_{n=1}^{\infty} E_n$

نوشته می‌شود و چون پیشامدهای E_n $(n \geq 1)$ نزولی هستند؛ در نتیجه بنا بر پیوستگی تابع احتمال داریم

$$P(\text{توپ شماره ۱ در ساعت ۱۲ نیمه‌شب در ظرف باشد})$$

$$= P\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n \to \infty} P(E_n) = \prod_{n=1}^{\infty} \frac{9n}{9n+1}$$

حال نشان می‌دهیم $\prod_{n=1}^{\infty} \frac{9n}{9n+1} = 0$ زیرا

$$\prod_{n=1}^{\infty} \frac{9n}{9n+1} = \left(\prod_{n=1}^{\infty} \frac{9n+1}{9n}\right)^{-1}$$

که معادل با این است که نشان دهیم $\prod_{n=1}^{\infty} \left(1 + \frac{1}{9n}\right) = \infty$. از طرفی به ازای هر $m \geq 1$ داریم

$$\prod_{n=1}^{\infty} \left(1 + \frac{1}{9n}\right) \geq \prod_{n=1}^{m} \left(1 + \frac{1}{9n}\right) = \left(1 + \frac{1}{9}\right)\left(1 + \frac{1}{18}\right)\left(1 + \frac{1}{27}\right) \cdots \left(1 + \frac{1}{9m}\right)$$

$$> \frac{1}{9} + \frac{1}{18} + \frac{1}{27} + \cdots + \frac{1}{9m} = \frac{1}{9} \sum_{i=1}^{m} \frac{1}{i}$$

حال با توجه به آنکه $\lim_{m \to \infty} \sum_{i=1}^{m} \frac{1}{i} = \infty$؛ پس

$$\prod_{m=1}^{\infty} \left(1 + \frac{1}{9n}\right) = \infty$$

در نتیجه اگر پیشامد F_1 به صورت توپ شماره ۱ ام در ساعت ۱۲ نیمه‌شب در ظرف باشد تعریف شود نشان داده‌ایم که $P(F_1) = 0$. به طور مشابه می‌توان نشان داد برای همه iها $P(F_i) = 0$ است و بنابراین احتمال اینکه ظرف در ساعت ۱۲ نیمه شب خالی نباشد، از رابطه $P\left(\bigcup_{i=1}^{\infty} F_i\right)$ به دست می‌آید که با استفاده از نامساوی بول داریم

$$P\left(\bigcup_{i=1}^{\infty} F_i\right) \leq \sum_{i=1}^{\infty} P(F_i) = 0$$

بنابراین با احتمال ۱، ظرف در ساعت ۱۲ خالی خواهد بود.

مقدمه: یکی از مهم ترین مفاهیم نظریه احتمال یعنی مفهوم احتمال شرطی را در اینجا مطرح می کنیم. این مفهوم اهمیت مضاعفی دارد، زیرا اولاً برای محاسبه احتمال پیشامدهایی استفاده می شود که قسمتی از اطلاعات آن در نتیجه انجام یک آزمایش در دسترس است. ثانیاً حتی اگر قسمتی از اطلاعات در دسترس نباشد، غالباً با به کار بردن احتمال شرطی به عنوان یک ابزار می توان محاسبه احتمال پیشامدها را ساده تر نمود.

مثال: یک تاس را که شانس رخداد عدد زوج در آن دو برابر عدد فرد می باشد را یک بار پرتاب می کنیم.

الف- احتمال رخداد یک عدد زوج را بیابید.

ب- اگر بدانیم عدد حاصل از پرتاب تاس از ۳ بزرگتر بوده است. احتمال رخداد یک عدد زوج را بیابید.

S	۱	۲	۳	۴	۵	۶
احتمالات	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$

$F = \{2, 4, 6\}$ $\qquad P(F) = \frac{6}{9} = \frac{2}{3}$

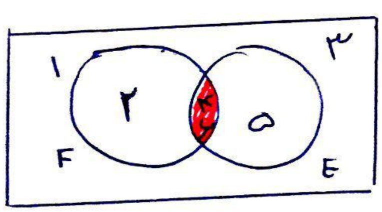
ب- براساس اطلاع داده شده فضای نمونه جدید

$E = \{4, 5, 6\}$

E	۴	۵	۶
احتمالات	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

$P(F|E) = \frac{4}{5}$

احتمال شرطی $P(F|E)$



$P(F|E) \propto P(E\cap F)$

$\Rightarrow P(F|E) = k\,P(E\cap F) \overset{\text{since}}{\Rightarrow} P(E|E)=1$

$\text{thus} \quad 1 = P(E|E) = k\,P(E\cap E) = k\,P(E) \overset{P(E)\neq 0}{\Rightarrow} k = \frac{1}{P(E)}$

$P(F|E) = \frac{P(F\cap E)}{P(E)} \qquad P(E)\neq 0$

در مثال فوق $P(E) = \frac{5}{9}$ و $P(E\cap F) = \frac{4}{9}$ بنابراین

$$P(F|E) = \frac{P(E\cap F)}{P(E)} = \frac{4/9}{5/9} = \frac{4}{5}$$

تعریف: احتمال شرطی پیشامد F به شرط وقوع پیشامد E، $P(F|E)$، عبارت است از

$$P(F|E) = \frac{P(E\cap F)}{P(E)} \qquad P(E) \neq 0$$

مثال: سکه‌ای را دو مرتبه پرتاب می‌کنیم. اگر فرض کنیم همه ۴ عضو فضای نمونه آزمایش $S = \{(H,H), (H,T), (T,H), (T,T)\}$ هم شانس باشند، احتمال شرطی اینکه هر دو پرتاب شیر ظاهر شوند به شرط اینکه پرتاب اول شیر آمده باشد چقدر است؟

حل:

$$E = \{(H,H)\} \qquad F = \{(H,H), (H,T)\}$$

$$P(E|F) = \frac{P(E\cap F)}{P(F)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$$

مثال: ظرفی شامل ۲۵ لامپ است که ۵ تای آن در شرایط خوب قابل استفاده‌اند برای حداقل ۳۰ روز کار خواهند کرد. ۱۰ لامپ کمی معیوب هستند و در مدت روز پس از استفاده خراب می‌شوند و ۱۰ لامپ باقی مانده کاملاً معیوب و غیر قابل استفاده‌اند. اگر لامپی را به تصادف از ظرف انتخاب کنیم و روشن شود احتمال اینکه این لامپ پس از یک هفته کار کند را بدست آورید.

G: پیشامد انتخاب تصادفی یک لامپ با شرایط خوب

T: پیشامد روشن نشدن لامپ

$$P(G|T^c) = \frac{P(G\cap T^c)}{P(T^c)} = \frac{P(G)}{P(T^c)} = \frac{\frac{5}{25}}{\frac{15}{25}} = \frac{1}{3}$$

باید توجه داشت که احتمال شرطی را با کاهش فضای نمونه آزمایش نیز می‌توان به دست آورد. زیرا وقتی که بدانیم لامپ روشن نمی‌شود آنگاه مسئله به انتخاب یک لامپ از ظرفی که دارای ۵ لامپ خوب و ۱۰ لامپ کمی معیوب است تبدیل می‌شود.

مثال: در ظرفی ۵۲ توپ از ۴ رنگ مختلف (قرمز، آبی، سبز و سفید) که هر رنگ با شماره‌های ۱ تا ۱۳ مشخص شده‌اند وجود دارد. این توپ‌ها را به تصادف و به طور مساوی بین ۴ نفر A، B، C و D تقسیم می‌کنیم. اگر افراد A و B جمعاً ۸ توپ قرمز دریافت کنند، احتمال اینکه نزد C از ۵ توپ قرمز باقی‌مانده ۳ توپ را دریافت کند چقدر است؟

حل: شاید ساده‌ترین راه این باشد که از فضای نمونه کاهش یافته استفاده کنیم یعنی با داشتن اطلاع اینکه افراد A و B از ۲۶ توپ دریافت شده ۸ توپ قرمز دارند، از ۲۶ توپ باقی‌مانده دقیقاً ۵ توپ قرمز وجود دارد که بین افراد C و D تقسیم شده است. چون هر توپ با شانس برابر توزیع می‌شود بنابراین احتمال اینکه نزد C، ۳ توپ قرمز بین ۱۳ توپ خود داشته باشد برابر است با

$$\frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} \approx 0.339$$

مثال: شرکته‌ای که در آن آقای x کار می‌کند یک مهمانی شام برای کارمندانی که حداقل یک پسر داشته باشند ترتیب داده است. اگر بدانیم که آقای x دو فرزند دارد و به مهمانی دعوت شده است، احتمال اینکه هر دو فرزند او پسر باشند چقدر است؟

فرض کنید که فضای نمونه آزمایش به صورت زیر باشد و همه نتایج دارای شانس برابر باشند

$$S = \{(b,b), (b,g), (g,b), (g,g)\}$$

حل : نشان دهنده اینکه حداقل یک فرزند آقای x پسر است $A:$

$B:$ پیشامد آنکه هر دو فرزند آقای x پسر باشد

$$P(B|A) = \frac{P(B\cap A)}{P(A)} = \frac{P(\{(b,b)\})}{P(\{b,b\},\{b,g\},\{g,b\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

بسیاری به طور نادرست دلیل می‌آورند که احتمال شرطی اینکه دو فرزند پسر باشند به شرط حداقل یک پسر، برابر با $\frac{1}{2}$ به جای مقدار صحیح $\frac{1}{3}$ است. زیرا آنها دلیل می‌آورند که فرزند دیگر آقای x که در مهمانی شرکت می‌کند با شانس برابر پسر یا دختر است و اشتباه آنها این است که فرض می‌کنند تنها این دو امکان وجود دارد و هم شانس هستند درحالی که از ابتدا چهار نتیجه هم شانس وجود دارد و اطلاع اینکه نتیجه (g,g) وجود ندارد ما را با سه پیشامد هم شانس $\{(b,b),(b,g),(g,b)\}$ روبرو می‌کند که نتیجتاً شانس اینکه فرزند دیگر آقای x دختر باشد دو برابر این است که او پسر باشد.

قانون ضرب احتمال : اگر E و F دو پیشامد باشند که بتوان همزمان اتفاق بیفتند آنگاه

$$P(E\cap F) = P(E)\,P(F|E)$$

مثال : دانشجویی مردد است که درس فرانسه و یا درس شیمی را انتخاب کند. او حدس می‌زند که با احتمال $\frac{1}{2}$ در درس فرانسه نمره A و با احتمال $\frac{2}{3}$ در درس شیمی نمره A خواهد آورد. اگر او تصمیم بگیرد که براساس نتیجه حاصل از پرتاب یک سکه درس را انتخاب کند مطلوب است احتمال اینکه وی در درس شیمی نمره A خواهد گرفت.

$A:$ پیشامد آنکه دانشجو نمره A در هر درسی که انتخاب می‌کند کسب کند

$C:$ پیشامد اینکه دانشجو درس شیمی را انتخاب کند

$$P(C\cap A) = P(C)P(A|C) = \frac{1}{2}\times\frac{2}{3} = \frac{1}{3}$$

مثال: فرض کنید ظرفی دارای ۸ توپ قرمز و ۴ توپ سفید است. دو توپ را بدون جایگذاری انتخاب می‌کنیم. اگر فرض کنیم که انتخاب هر یک از توپ هم شانس است، احتمال اینکه هر دو توپ انتخابی قرمز باشد چقدر است؟

R_i: پیشامد آنکه توپ iام قرمز باشد

$i = 1, 2$

$$P(R_1\cap R_2) = P(R_1)P(R_2|R_1) = \frac{8}{12}\times\frac{7}{11} = \frac{14}{33}$$

احتمال فوق را می‌توان به صورت زیر نیز به دست آورد

$$P(R_1\cap R_2) = \frac{\binom{8}{2}}{\binom{12}{2}}$$

قانون ضرب احتمال در حالت کلی برای n پیشامد را می‌توان به صورت زیر نوشت

$$P(E_1\cap E_2\cap\dots\cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1\cap E_2)\times$$

$$\dots\times P(E_n|E_1\cap E_2\cap\dots\cap E_{n-1})$$

زیرا طرف راست تساوی فوق برابر است با:

$$P(E_1)\frac{P(E_2\cap E_1)}{P(E_1)}\times\frac{P(E_1\cap E_2\cap E_3)}{P(E_1\cap E_2)}\times\dots\times\frac{P(E_1\cap E_2\cap\dots\cap E_n)}{P(E_1\cap\dots\cap E_{n-1})}$$

$$= P(E_1\cap E_2\cap\dots\cap E_n)$$

مثال: در فصل قبل مثالی عنوان شد بدین صورت که: در ظرفی ۱۲ توپ از ۴ رنگ مختلف قرمز، آبی، سبز و سفید وجود دارد. احتمال اینکه هر یک از ۴ نفر توپ شماره ارا دریافت کنند را با استفاده از قانون ضرب احتمال محاسبه کنید.

E_1: توپ شماره ۱ آبی نزد یکی از افراد باشد

E_2: توپ شماره ۱ آبی و توپ شماره ۱ قرمز نزد دو نفر متفاوت باشند

E_3: توپ شماره ۱ آبی، توپ شماره ۱ قرمز و توپ شماره ۱ سبز نزد افراد متفاوت باشند

E_4: همه توپ‌ها با شماره یک نزد افراد متفاوت باشند

$$P(E_1\cap E_2\cap E_3\cap E_4)=P(E_1)P(E_2|E_1)P(E_3|E_1\cap E_2)$$
$$\times P(E_4|E_1\cap E_2\cap E_3)$$

چون E_1 فضای نمونه آزمایش است پس $P(E_1)=1$ ، از طرفی چون که توپ شماره ۱ آبی را داشته باشد ۱۲ توپ از ۵۱ توپ دیگر را خواهد داشت. بنابراین

$$P(E_2|E_1)=\frac{39}{51}$$

همچنین از آنجایی که توپ شماره ۱ آبی و توپ شماره ۱ قرمز را داشته باشند ۲۴ توپ دیگر از ۵۰ توپ باقی مانده را خواهند داشت. بنابراین

$$P(E_3|E_1\cap E_2)=\frac{26}{50}$$

$$P(E_4|E_1\cap E_2\cap E_3)=\frac{13}{49}$$

$$\Rightarrow P(E_1\cap E_2\cap E_3\cap E_4)=\frac{39\times 26\times 13}{51\times 50\times 49}\approx 0/105$$

پس تقریباً ۱۰٫۵ درصد شانس دارد که هر تیم یک توپ با شماره ۱ داشته باشد

تذکر: تعریف ارائه شده از $P(E|F)$ ، با تفسیر احتمال به عنوان فراوانی نسبی با تکرار زیاد، سازگار است. برای دیدن این واقعیت فرض کنید که آزمایش n مرتبه تکرار می‌شود وقتی که n عدد بزرگی است. در این صورت اگر فقط آزمایش‌هایی را در نظر بگیریم که F اتفاق افتاده است، آنگاه $P(E|F)$ برابر است با نسبت آنهایی که E هم اتفاق افتاده باشد. برای اثبات توجه کنید که چون $P(F)$ نسبت وقوع پیشامد F است پس تعداد دفعاتی که در n آزمایش F اتفاق می‌افتد برابر با $nP(F)$ بوده و به طریق مشابه $nP(E\cap F)$ تعداد دفعاتی است که E و F باهم اتفاق می‌افتند و بنابراین از بین تقریباً $nP(F)$ آزمایش که F اتفاق

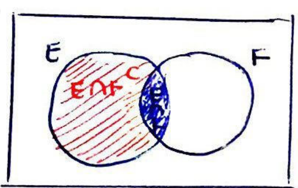
افتاده، نسبت آنهایی که E نیز اتفاق افتاده تقریباً برابر است با

$$\frac{n\,P(E\cap F)}{n\,P(F)} = \frac{P(E\cap F)}{P(F)}$$

این تقریب دقیق‌تر خواهد شد هرگاه n بزرگتر و بزرگتر شود و می‌بینیم که تعریف مناسب برای $P(E|F)$ داریم.

قضیه احتمال کل:

اگر E و F دو پیشامد باشند می‌توانیم پیشامد E را به صورت $E=(E\cap F)\cup(E\cap F^c)$ در نظر بگیریم، زیرا برای آنکه نقطه‌ای در E باشد باید یا در هر دو E و F باشد، یا آنکه در E باشد و در F نباشد. چون $E\cap F$ و $E\cap F^c$ ناسازگار هستند بر اساس اصل ۳ داریم



$$P(E) = P(E\cap F)+P(E\cap F^c) = P(E|F)\,P(F) + P(E|F^c)\,P(F^c) = P(E|F)\,P(F) + P(E|F^c)\,[1-P(F)]$$

تفکیک فضای نمونه به چند پیشامد

گوییم فضای نمونه S توسط پیشامدهای F_1، F_2، ...، F_k افراز شده هرگاه برای هر $i=1,\dots,k$، $F_i\neq\varnothing$ و برای هر k، $i\neq j=1,2,\dots,k$، $F_i\cap F_j=\varnothing$ و نیز $\bigcup_{i=1}^{k}F_i=S$.

اکنون می‌خواهیم قضیه احتمال کل را در حالت کلی بیان کنیم:

فرض کنیم فضای نمونه S به پیشامدهای F_1، F_2، ...، F_k تفکیک شده. برای هر پیشامد دیگر E که دارای احتمال مثبت باشد داریم

$$E = E\cap S = E\cap\left(\bigcup_{i=1}^{k}F_i\right) = \bigcup_{i=1}^{k}(F_i\cap E)$$

چون $(E\cap F_i)\cap(E\cap F_j)=\varnothing$ لذا

$$P(E) = P\left(\bigcup_{i=1}^{k} (F_i \cap E)\right) = \sum_{i=1}^{k} P(E_i \cap E) = \sum_{i=1}^{k} P(E|F_i)P(F_i)$$

پس بنابراین

$$P(E) = \sum_{i=1}^{k} P(E|F_i)P(F_i)$$

مثال : (قسمت اول) یک شرکت بیمه اعتقاد دارد که مردم را می‌توان به دو گروه مستعد برای تصادف و غیر مستعد برای تصادف تقسیم بندی شود. آمار این شرکت نشان می‌دهد که یک فرد مستعد تصادف در یک سال با احتمال ۰/۴ تصادف خواهد داشت و این احتمال برای فرد غیر مستعد تصادف به ۰/۲ کاهش می‌یابد. اگر فرض کنیم ۳۰ درصد از افراد جامعه مستعد تصادف هستند. احتمال این که یک فرد بیمه شده جدید در یک سال تصادف داشته باشد را به دست آورید.

A_1 : پیشامد اینکه فرد بیمه شده در یک سال تصادف داشته باشد

A : پیشامد اینکه فرد مستعد تصادف باشد

$$P(A_1) = P(A_1|A)P(A) + P(A_1|A^c)P(A^c) = 0/4 \times 0/3 + 0/2 \times 0/7 = 0/26$$

مثال : (قسمت دوم) فرض کنید یک فرد بیمه شده در طول یک سال تصادف کرده باشد، احتمال اینکه او از گروه مستعد تصادف باشد چقدر است؟

$$P(A|A_1) = \frac{P(A\cap A_1)}{P(A_1)} = \frac{P(A)P(A_1|A)}{P(A_1)} = \frac{(0/3)(0/4)}{0/26} = \frac{6}{13}$$

مثال: در ظرفی ۵۲ توپ از ۴ رنگ مختلف (قرمز، آبی، سبز و سفید) که هر رنگ با شماره‌های ۱ تا ۱۳ مشخص شده‌اند وجود دارد. از این ظرف توپ را یک به یک خارج می‌کنیم. قبل از خارج کردن توپ بعدی حدس بزنند که توپ بعدی توپ شماره ۱ قرمز است برنده می‌شود. (حتی اگر فقط یک توپ در ظرف باقی مانده باشد و توپ شماره ۱ قرمز خارج نشده باشد)

به نظر شما روش حدس زدن درست و نادرست در این بازی چیست؟

حل: هر روشی که برای انجام این بازی به کار برده شود دارای احتمال برد $\frac{1}{52}$ است. برای اثبات این موضوع از روش استقراء برای n توپ استفاده می‌کنیم و نشان می‌دهیم که مستقل از هر روشی که برای حدس زدن به کار بریم احتمال برد $\frac{1}{n}$ است. واضح است که این نتیجه برای $n=1$ صادق است. پس اگر برای $n-1$ توپ صحیح باشد، فرض با n توپ را در نظر می‌گیریم (یکی از توپها، توپ شماره ۱ قرمز است). با هر روش حدس زدن اگر اولین انتخاب را توپ شماره ۱ قرمز حدس بزنیم، احتمال برد $\frac{1}{n}=p$ است. از طرف دیگر اگر انتخاب اول را توپ شماره ۱ قرمز حدس نزنیم پس احتمال برد برابر است با احتمال اینکه اولین توپ قرمز نباشد یعنی $\frac{n-1}{n}$ ضرب در احتمال شرطی برنده شدن به شرط آنکه اولین انتخاب، توپ شماره ۱ قرمز نباشد. اما این احتمال شرطی برابر است با احتمال برد در حالتی که $n-1$ توپ داریم که شامل توپ شماره ۱ قرمز است یعنی $\frac{1}{n-1}$ پس احتمال برد به شرط اینکه اولین انتخاب را حدس نزنیم برابر است با: $\frac{n-1}{n}\times\frac{1}{n-1}=\frac{1}{n}$

بنابراین اگر G بیانگر حدس زدن اولین توپ باشد، داریم

$$P(\text{برد}) = P(\text{برد}\mid G)P(G) + P(\text{برد}\mid G^c)P(G^c)$$
$$= \frac{1}{n}p + \frac{1}{n}(1-p) = \frac{1}{n}$$

مثال: در پاسخ به یک آزمون چندجوابی، دانشجویی یا پاسخ سوال را می‌داند و یا آن را حدس می‌زند. اگر p احتمال این باشد که دانشجو پاسخ سوال را بداند و $1-p$ احتمال اینکه دانشجو حدس می‌زند. فرض کنید دانشجویی که حدس می‌زند با احتمال $\frac{1}{m}$ پاسخ صحیح بدهد وقتی که m تعداد پاسخهای هر سوال باشد. مطلوب است احتمال شرطی اینکه دانشجو پاسخ یک سوال را می‌دانسته به شرط اینکه پاسخ صحیح داده باشد؟

C : پیشامد پاسخ صحیح دادن به سؤال

K : پیشامد دانستن پاسخ صحیح

$$P(K|C) = \frac{P(C|K)P(K)}{P(C|K)P(K)+P(C|K^c)P(K^c)} = \frac{p}{p+\frac{1}{m}(1-p)} = \frac{mp}{1+(m-1)p}$$

مثلاً اگر $m=5$ و $p=\frac{1}{2}$ باشد، آنگاه احتمال اینکه دانشجو پاسخ صحیح سؤال را می‌دانسته به شرط اینکه پاسخ سوال را صحیح داده باشد برابر $\frac{5}{6}$ است.

مثال: نوعی آزمایش خون برای تشخیص یک بیماری، وقتی که بیماری وجود دارد در ۹۵٪ مواقع مؤثر است. همچنین نتیجه آزمایش می‌تواند اشتباهاً بیان کننده وجود بیماری در ۱٪ از اوقات برای فرد سالم باشد (یعنی اگر خون یک فرد سالم آزمایش شود با احتمال ۰/۰۱ نتیجه این خواهد بود که وی بیمار است.) اگر ۰/۵٪ از افراد جامعه واقعاً این بیماری را داشته باشند. احتمال اینکه فردی که نتیجه آزمایش خون او مثبت است دارای بیماری باشد را به دست آورید.

حل:

I : نشان دهنده پیشامد اینکه فرد آزمایش شده بیمار است

M : پیشامد اینکه نتیجه آزمایش مثبت است

$$P(I|M) = \frac{P(I\cap M)}{P(M)} = \frac{P(M|I)P(I)}{P(M|I)P(I)+P(M|I^c)P(I^c)}$$

$$= \frac{0/95 \times 0/005}{0/95\times 0/005 + 0/01 \times 0/995} = \frac{95}{294} \approx 0/323$$

بنابراین فقط ۳۲ درصد از افرادی که نتیجه آزمایش آنها مثبت است واقعاً بیمار هستند.

از آنجایی که بسیاری از دانشجویان از به دست آمدن چنین نتیجه‌ای متعجب می‌شوند (زیرا آنها انتظار دارند که نتیجه عدد بزرگتری باشد چون آزمایش توانایی به نظر می‌رسد)

بهتر است از روش دیگر که ساده تر و آشکارتر است مسأله را حل کنیم .

چون ۰/۵ درصد از افراد جامعه واقعاً بیمار هستند بنا براین به طور متوسط از هر ۲۰۰ نفری که مورد آزمایش قرار می‌گیرند یک نفر بیمار است و آزمایش با احتمال ۹۵٪ خواهد گفت که او بیمار است . در نتیجه به طور متوسط از هر ۲۰۰ نفر آزمایش شده ، نتیجه آزمایش به طور صحیح خواهد گفت که ۰/۹۵ فرد بیمار است . از طرف دیگر از ۱۹۹ نفر سالم ، آزمایش اشتباهاً می‌گوید که (۰/۰۱)(۱۹۹) افراد بیمار هستند . در مجموع برای هر ۰/۹۵ فرد بیمار آزمایش به طور صحیح بیان می‌دارد که او بیمار است و برای هر (۰/۰۱)(۱۹۹) فرد سالم آزمایش اشتباهاً بیان می‌دارد که او بیمار است . در نتیجه نسبت دفعاتی که نتیجه صحیح است وقتی که آزمایش می‌گوید فرد بیمار است برابر است با

$$\frac{۰/۹۵}{۰/۹۵ + (۱۹۹)(۰/۰۱)} = \frac{۹۵}{۲۹۴} \approx ۰/۳۲۳$$

مثال : پزشکی را در نظر بگیرید که براساس روش زیر طبابت می‌کند .

"اگر حداقل ۸۰٪ مطمئن باشد که بیمار دارای بیماری خاصی است پیشنهاد عمل جراحی را می‌دهد و اگر کاملاً مطمئن نباشد پیشنهاد یک آزمایش دیگر که نسبتاً گران و بعضاً دردآور است داده می‌شود ، او ۶۰ درصد اطمینان دارد که آقای x بیمار است . بنا براین دستور انجام آزمایشهای سری A را می‌دهد که چنانچه بیماری وجود داشته باشد همیشه پاسخ آن مثبت است و اگر وی بیمار نباشد تقریباً هیچ وقت نتیجه مثبت نخواهد داشت . نتیجه آزمایشها مثبت بود و پزشک هم مقدمات انجام عمل جراحی را فراهم نمود تا اینکه آقای x اطلاع داد که او بیماری قند دارد . این اطلاع ، موضوع را پیچیده کرد زیرا اگر چه اطمینان ۶۰ درصد اولیه را تغییر نداد ، اما در تفسیر نتیجه آزمایشهای A تأثیر گذاشت . زیرا آزمایشهای گروه A اگر بیمار دارای بیماری قند باشد در ۳۰ درصد اوقات به

خطا نتیجه مثبت می‌دهد. با وجود این شرایط چه باید کرد؟ آیا آزمایش دیگری داده شود یا

بلافاصله بیمار عمل شود؟»

حل: برای اینکه تصمیم گرفته شود که بیمار عمل شود یا نه، ابتدا بایستی احتمال اینکه وی بیمار

باشد را با توجه به اینکه نتیجه آزمایش سری A مثبت است محاسبه کرد.

D : پیشامد آنکه آقای x بیمار است

E : پیشامد آنکه نتیجه آزمایش‌های سری A مثبت باشد

$$P(D|E) = \frac{P(D\cap E)}{P(E)} = \frac{P(D)\,P(E|D)}{P(E|D)\,P(D) + P(E|D^c)\,P(D^c)} = \frac{(0/6)\,1}{1(0/6) + (0/3)(0/4)} = 0/833$$

چون دکتر بیش از ۸۰ درصد مطمئن است که آقای x بیمار است پس باید او را مورد عمل جراحی

قرار دهد.

مثال: در یک مرحله پس از محاکمه جنایی، بازپرس به اندازه ۶۰ درصد از گناهکار بودن یک متهم

اطمینان دارد. حال فرض کنید که براساس مدرک جدید معلوم شود که جنایتکار دارای ویژگی

خاصی است (مثلاً چپ دست، کم مو یا قهوه‌ای)؛ اگر ۲۰ درصد از افراد جامعه دارای

این ویژگی باشند و متهم نیز این ویژگی را داشته باشد بازپرس چقدر نسبت به گناهکار بودن

او اطمینان پیدا می‌کند.

C : پیشامد آنکه متهم دارای ویژگی گناهکاران باشد یعنی چپ دست، کم مو یا قهوه‌ای باشد

G : پیشامد آنکه متهم گناهکار است

$$P(G|C) = \frac{P(G\cap C)}{P(C)} = \frac{P(C|G)\,P(G)}{P(C|G)\,P(G) + P(C|G^c)\,P(G^c)} = \frac{1\,(0/6)}{1(0/6) + (0/2)(0/4)} \approx 0/882$$

در اینجا فرض کردیم احتمال اینکه متهم ویژگی را دارا باشد، درصورتی که واقعاً بی‌گناه باشد برابر ۰/۲ است

که برابر نسبت افرادی است که آن ویژگی را دارند.

مثال: در مسابقات جهانی یک بازی شانس (بریج) که در شهر بوینس آیرس در ماه مه ۱۹۶۵ برگزار شد، دو همبازی به تقلب در بازی متهم گردیدند، بدین صورت که آنها به وسیله استفاده از انگشتان اطلاعاتی را در اختیار یکدیگر قرار می‌دادند. برای بررسی این موضوع، دادگاهی تشکیل شد که در طول دادگاه تمامی بازیهای جدیدی را که به وسیله آنها انجام گرفت بررسی نموده و ادعا کرد که این بازیها با فرض تقلب کردن سازگار است و آنها را گناهکار تشخیص داده. در این رابطه وکیل مدافع نیز اعلام نمود که این بازیها کاملاً عادی است. نهایتاً هیئت منصفه دادگاه رأی داد تا زمانی که بازی آنها با فرض گناهکار بودن سازگار باشد باید آنها را مجرم داشت. نظر شما درباره این قضاوت چیست؟

H: نشان دهنده یک فرض معین (مانند مجرم بودن دو همبازی)

E: نشان دهنده شاهد جدید

$$P(H|E) = \frac{P(H\cap E)}{P(E)} = \frac{P(E|H)\,P(H)}{P(E|H)\,P(H) + P(E|H^c)\,[1-P(H)]} \qquad (*)$$

که $P(H)$ ارزیابی ما از احتمال صحت فرض قبل از ظهور شاهد جدید است. شاهد جدید پشتیبان فرض است وقتی که آن را محتمل‌تر کند، یعنی وقتی که $P(H|E) \geq P(H)$

بنابراین با توجه به $(*)$ داریم

$$P(E|H) \geq P(E|H)\,P(H) + P(E|H^c)\,[1-P(H)]$$

$$\Rightarrow P(E|H) \geq P(E|H^c)$$

به عبارت دیگر یک شاهد جدید را می‌توان پشتیبان صحت یک فرض دانست فقط اگر وقوع آن محتمل‌تر باشد وقتی که فرض صحیح است نسبت به زمانی که غلط است. در واقع احتمال صحت فرض بستگی به احتمال اولیه آن و نسبت احتمالهای شرطی دارد، زیرا

از رابطه $(*)$ داریم

$$P(H|E) = \frac{P(H)}{P(H) + [1-P(H)]\,\dfrac{P(E|H^c)}{P(E|H)}}$$

بنابراین در مسئله تحت بررسی بازیکران نمی‌توانند به عنوان حمایت فرض تقلب کردن در نظر گرفته شود فقط اگر احتمال برد این بازی در صورت تقلب، بیشتر از حالت عدم تقلب باشد. اما چون شاهد چنین ادعایی را نکرد بنابراین اعلام اینکه شاهد در جهت حمایت فرض گناهکار بودن آنها است بی‌ارزش می‌باشد.

تعیین در احتمال یک فرض وقتی که یک شاهد جدید معرفی می‌شود را می‌توان به طور خلاصه به عنوان تغییر در نسبت بخت یک فرض بیان نمود.

تعریف : نسبت بخت پیشامد A عبارتست از

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)}$$

به عبارت دیگر نسبت بخت پیشامد A می‌گوید که چه نسبتی بین شانس اتفاق افتادن پیشامد A و اتفاق نیفتادن آن وجود دارد. برای مثال اگر $P(A) = \frac{2}{3}$ باشد آنگاه $P(A) = 2P(A^c)$ بنابراین نسبت بخت برای پیشامد A برابر با ۲ است. اگر نسبت بخت درستی یک فرض برابر با α باشد آنگاه بیان عبارت شانس درست بودن فرض «α به یک» است بسیار رایج می‌باشد.

حال فرض H را که احتمال درستی آن $P(H)$ است در نظر بگیرید، چنانچه شاهد E معرفی شود آنگاه احتمال‌های مشروط درستی H و نادرستی آن برابر است با

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)}$$

$$P(H^c|E) = \frac{P(E|H^c)\,P(H^c)}{P(E)}$$

بنابراین نسبت بخت جدید بعد از پیشامد E برابر است با

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)\,P(E|H)}{P(H^c)\,P(E|H^c)}$$

یعنی مقدار جدید نسبت بخت H برابر است با مقدار قدیم این نسبت ضرب در نسبت احتمال شاهد E به شرط آنکه H درست باشد به احتمال شاهد E به شرط آنکه H نادرست باشد. این مطلب تأییدی بر نتیجه مثال محکمه جنایی است، زیرا نسبت بخت و به پیرو آن احتمال درستی H صعودی است وقتی که وقوع پیشامد شاهد در صورت درستی H محتمل‌تر از وقوع آن در صورت نادرستی H باشد. به همین ترتیب نسبت بخت نزولی است اگر وقوع پیشامد شاهد در صورت نادرستی H محتمل‌تر از وقوع آن در صورت درستی H باشد.

مثال: وقتی که سکه A را پرتاب می‌کنیم با احتمال $\frac{1}{4}$ شیر ظاهر می‌شود در حالی که احتمال ظاهر شدن شیر در سکه B برابر با $\frac{3}{4}$ است. فرض کنید یکی از این دو سکه را به تصادف انتخاب نموده و دو مرتبه آن را پرتاب می‌کنیم. اگر هر دو مرتبه شیر ظاهر شود احتمال اینکه سکه B پرتاب شده باشد را به دست آورید.

B: پیشامد پرتاب سکه B

$$P(B) = P(B^c)$$

$$\frac{P(B|HH)}{P(B^c|HH)} = \frac{9/16}{1/16} = 9$$

در نتیجه نسبت بخت ۹ به ۱ است. به تعبیر دیگر احتمال اینکه سکه B انتخاب و پرتاب شده باشد $\frac{9}{10}$ است.

قانون بیز

فرض کنید پیشامد E اتفاق افتاده باشد و بخواهیم احتمال اینکه یکی از پیشامدهای F_i ام اتفاق افتاده باشد را به دست آوریم

$$P(F_i|E) = \frac{P(E|F_i)P(F_i)}{\sum_{j=1}^{k} P(E|F_j)P(F_j)}$$

به هر کدام از پیشامدهای F_i پیشامدهای پیشین و به پیشامد $(F_i|E)$ پیشامد پسین می‌گویند.

مثال: هواپیمایی در مسیر پرواز خود ناپدید شده است و با شانس مساوی تصور می‌شود که در یکی از ۳ منطقه ممکن سقوط کرده است. فرض کنید $1-\beta_i$ نشان دهنده احتمال این باشد که هواپیما با جستجو در منطقه i ام وقتی که واقعاً در آن منطقه سقوط کرده است پیدا شود. (مقادیر ثابت β_i را احتمال‌های از بالا نگاه کردن می‌نامند و معمولاً براساس ویژگی‌های جغرافیایی و محیطی منطقه تعیین می‌شوند.) احتمال شرطی اینکه هواپیما در منطقه i ام سقوط کرده باشد به شرط اینکه نتیجه جستجو در منطقه ۱ ناموفق بوده است را بدست آورید $(i=1,2,3)$

R_i: پیشامد سقوط هواپیما در منطقه i

E: پیشامد اینکه جستجو در منطقه ۱ ناموفق بوده است

$$P(R_1|E)=\frac{P(E|R_1)\,P(R_1)}{\sum_{i=1}^{3}P(E|R_i)\,P(R_i)}=\frac{\beta_1\times\frac{1}{3}}{\beta_1\times\frac{1}{3}+1\left(\frac{1}{3}\right)+1\times\frac{1}{3}}=\frac{\beta_1}{\beta_1+2}$$

$$j=2,3\qquad P(R_j|E)=\frac{P(E|R_j)\,P(R_j)}{P(E)}=\frac{1\left(\frac{1}{3}\right)}{\beta_1\left(\frac{1}{3}\right)+\frac{1}{3}+\frac{1}{3}}=\frac{1}{\beta_1+2}$$

احتمال شرطی اینکه هواپیما در منطقه ۱ سقوط کرده باشد به شرط اینکه جستجوی آن منطقه ناموفق بوده، تابعی صعودی از β_1 است، یعنی هرچه β_1 بزرگتر باشد ناموفق بودن جستجو در پیدا کردن هواپیما به بدشانسی نسبت داده می‌شود، تا اینکه بگوییم هواپیما در آنجا نیست. به همین ترتیب $P(R_j|E)$ $(j\neq 1)$ تابعی نزولی از β_1 است.

مثال: فرض کنید ۳ کارت یکسان داریم که یکی از آنها هر دو طرفش قرمز، دیگری هر دو طرفش سیاه و کارت سوم یک طرفش قرمز و طرف دیگر آن سیاه است. سه کارت را مخلوط نموده و یکی از آنها را به تصادف انتخاب و روی زمین قرار می‌دهیم اگر طرف بالای کارت انتخاب شده قرمز باشد احتمال اینکه طرف قرار گرفته روی زمین سیاه باشد چقدر است؟

R: پیشامد آنکه طرف بالای کارت انتخاب شده قرمز است

RR: پیشامد آنکه کارت انتخاب شده هر دو طرفش قرمز

BB: پیشامد آنکه کارت انتخاب شده هر دو طرفش سیاه

RB: پیشامد آنکه کارت انتخاب شده یک طرفش قرمز و طرف دیگر آن سیاه است

$$P(RB|R) = \frac{P(R|RB)\,P(RB)}{P(R|RB)\,P(RB) + P(R|RR)\,P(RR) + P(R|BB)\,P(BB)}$$

$$= \frac{(\frac{1}{2})(\frac{1}{3})}{(\frac{1}{2})(\frac{1}{3}) + (1)(\frac{1}{3}) + 0 \times \frac{1}{3}} = \frac{1}{3}$$

یعنی پاسخ برابر با $\frac{1}{3}$ است. درحالی که بعضی از دانشجویان حدس می‌زنند که پاسخ $\frac{1}{2}$ صحیح است زیرا تصور می‌کنند که با معلوم شدن یک طرف کارت، دو حالت هم‌شانس وجود دارد که کارت‌های دو طرف قرمز و یا یک طرف قرمز و یک طرف سیاه هستند. اشتباه آنها در این است که تصور کنند این دو حالت هم‌شانس هستند، زیرا اگر فکر کنیم که هر کارت شامل دو طرف متفاوت باشد، آنگاه ۶ نتیجه هم‌شانس R_1، R_2، B_1، B_2 و R_3، B_3 می‌تواند وجود داشته باشد که نتیجه R_1 این است که یک طرف کارت دو طرف قرمز مشاهده شود و R_3 نتیجه این است که طرف قرمز کارت یک طرف قرمز و یک طرف سیاه دیده می‌شود و ...

از آنجایی که زمانی طرف دیگر کارت سیاه است که R_3 اتفاق افتاده باشد. بنابراین احتمال مورد نظر احتمال شرطی R_3 به شرط وقوع R_1 یا R_2 یا R_3 است که به وضوح برابر با $\frac{1}{3}$ می‌باشد.

مثال: خانواده‌ای با دو فرزند را درنظر بگیرید. فرض کنید به مادر خانواده که همراه یکی از بچه‌هایش قدم می‌زند برخورد نماییم و مشاهده کنیم که فرزند همراه او دختر است.

احتمال اینکه هر دو فرزند خانواده دختر باشند چقدر است؟

G_1 : بپیشامد اینکه اولین فرزند (فرزند بزرگتر) دختر است

G_2 : دومین فرزند دختر است

G : فرزندی که با مادرش مشاهده شده دختر است

B_1 : پیشامد اینکه اولین فرزند پسر است

B_2 : پیشامد اینکه فرزند دوم پسر است

$$P(G_1 \cap G_2 \mid G) = \frac{P(G_1 \cap G_2 \cap G)}{P(G)} = \frac{P(G_1 \cap G_2)}{P(G)}$$

$$= \frac{P(G_1 \cap G_2)}{P(G \mid G_1 \cap G_2)P(G_1 \cap G_2) + P(G \mid G_1 \cap B_2)P(G_1 \cap B_2) + P(G \mid B_1 \cap G_2)P(B_1 \cap G_2) + P(G \mid B_1 \cap B_2)P(B_1 \cap B_2)}$$

$$= \frac{P(G_1 \cap G_2)}{P(G_1 \cap G_2) + P(G \mid G_1 \cap B_2)P(G_1 \cap B_2) + P(G \mid B_1 \cap G_2)P(B_1 \cap G_2)}$$

حال اگر فرض کنیم که کلیه حالتهای پسر و دختر بودن فرزندان هم شانس و برابر با $\frac{1}{4}$ باشد داریم

$$P(G_1 \cap G_2 \mid G) = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{P(G \mid G_1 \cap B_2)}{4} + \frac{P(G \mid B_1 \cap G_2)}{4}}$$

$$= \frac{1}{1 + P(G \mid G_1 \cap B_2) + P(G \mid B_1 \cap G_2)}$$

بنابراین پاسخ سؤال بستگی به فرضیه‌ای دارد که در رابطه با فرزند مشخص شده با مادرش در نظر گرفته می‌شود. یعنی احتمالهای مشروط اینکه فرزند همراه مادر یک دختر است به شرط پیشامد $G_1 \cap B_2$ و همچنین فرزند همراه مادر یک دختر است به شرط $G_2 \cap B_1$.

برای مثال اگر بخواهیم مستقل از جنس بچه‌ها فرض کنیم فرزند همراه مادر با احتمال P فرزند بزرگتر است داریم

$$P(G \mid G_1 \cap B_2) = P = 1 - P(G \mid B_1 \cap G_2)$$

$$\Rightarrow P(G_1 \cap G_2 \mid G) = \frac{1}{2}$$

از طرف دیگر اگر فرض کنیم که وقتی بچه ها از جنسهای متفاوت هستند مادر با احتمال p دخترش را مستقل از سن او همراه خود می برد داریم

$$P(G \mid G_1 \cap B_2) = P(G \mid B_1 \cap G_2) = p$$

$$\Rightarrow P(G_1 \cap G_2 \mid G) = \frac{1}{1+2p}$$

برای مثال اگر $p=1$ یعنی مادر همیشه با دخترش قدم می زند آنگاه احتمال داشتن دو دختر برابر با $\frac{1}{3}$ است که با مثالی که قبلاً به آن پرداخته شد مطابقت دارد، زیرا دیدن مادر همراه دخترش معادل این است که آنها حداقل یک دختر دارند.

مثال: در ظرفی ۳ نوع فلاش دوربین عکاسی وجود دارد. احتمال اینکه فلاش نوع ۱ بیش از ۱۰۰ ساعت کار کند برابر با ۰/۷ و این احتمال برای فلاشهای نوع ۲ و ۳ به ترتیب برابر ۰/۴ و ۰/۳ می باشد. فرض کنید ۲۰ درصد از فلاشهای موجود در ظرف از نوع ۱، ۳۰ درصد از نوع ۲ و ۵۰ درصد از نوع ۳ باشد.

الف - احتمال اینکه یک فلاش از ظرف انتخاب شود و بیش از ۱۰۰ ساعت کار کند را به دست آورید.

ب - اگر فلاش انتخاب شده بیش از ۱۰۰ ساعت کار کند احتمال شرطی اینکه این فلاش از نوع j باشد چقدر است؟ $j=1,2,3$

A: پیشامد آنکه فلاش بیش از ۱۰۰ ساعت کار کند

F_j: پیشامد آنکه فلاش نوع j، $(j=1,2,3)$ انتخاب شود

$$P(A) = P(A \mid F_1)P(F_1) + P(A \mid F_2)P(F_2) + P(A \mid F_3)P(F_3)$$

$$= (0/7)(0/2) + (0/4)(0/3) + (0/3)(0/5) = 0/41$$

یعنی ۴۱٪ شانس دارد که فلاش انتخاب شده بیش از ۱۰۰ ساعت کار کند.

ب -

$$P(F_j|A) = \frac{P(A|F_j)P(F_j)}{P(A)}$$

$$P(F_1|A) = \frac{(0/8)(0/2)}{0/41} = \frac{16}{41} \qquad P(F_2|A) = \frac{(0/4)(0/3)}{0/41} = \frac{12}{41}$$

$$P(F_3|A) = \frac{(0/3)(0/5)}{0/41} = \frac{15}{41}$$

بنابراین دقت شود، احتمال اولیه انتخاب فلاش نوع ۱ برابر با ۰/۲ است ولی با داشتن اینکه فلاش بیش از ۱۰۰ ساعت کارکرده این احتمال به $0/39 \simeq \frac{16}{41}$ افزایش می‌یابد.

مثال: یک شرکت تولیدی مواد اولیه مورد نیاز خود را تنها از دو کارخانه A و B تهیه می‌کند. این کارخانه‌ها به ترتیب ۱۲ و ۸ درصد محصولاتشان معیوب است. مسئول انبار این شرکت تولیدی می‌داند که اگر کالای معیوبی وجود داشته باشد در $\frac{1}{3}$ مواقع این کالا تولید شده در کارخانه A و در $\frac{2}{3}$ مواقع تولید شده در کارخانه B بوده است. معین کنید که این شرکت چند درصد خرید مواد اولیه خود را از شرکت A انجام می‌دهد؟

D: پیشامد معیوب بودن کالا

$$P(A|D) = \frac{P(A)P(D|A)}{P(D|A)P(A) + P(D|B)P(B)} \Rightarrow \frac{1}{3} = \frac{P(A) \times 0/12}{P(A) \times 0/12 + P(B) \times 0/08}$$

حال با توجه به آنکه $P(B) = 1 - P(A)$

$$0/36P(A) = 0/12P(A) + 0/08(1 - P(A))$$

$$\Rightarrow 0/32P(A) = 0/08 \Rightarrow P(A) = \frac{1}{4}$$

پیشامدهای مستقل

E مستقل از F است هرگاه اطلاع از وقوع پیشامد F احتمال وقوع پیشامد E را تغییر ندهد.

حال چون $P(E|F)=\frac{P(E\cap F)}{P(F)}$ پس E و F مستقل اند اگر

$$P(E\cap F)=P(E)P(F)$$

از آنجائیکه رابطه فوق برای E و F متقارن است (یعنی با جابجا کردن E و F تغییری در رابطه ایجاد نمی‌شود) لذا اگر E مستقل از F باشد، F نیز مستقل از E است. دو پیشامد E و F که مستقل نباشند را وابسته گویند.

مثال: فرض کنید ۵۲ توپ از ۴ رنگ مختلف (قرمز، آبی، سبز و سفید) که هر رنگ با شماره‌های ۱ تا ۱۳ مشخص شده‌اند وجود دارد. یک توپ به تصادف از این ظرف انتخاب می‌کنیم. اگر E نشان دهنده پیشامد اینکه توپ شماره ۱ انتخاب شده باشد و F نشان دهنده پیشامد قرمز بودن توپ انتخاب شده باشد آنگاه E و F مستقلند، زیرا

$$P(E\cap F)=\frac{1}{52} \qquad P(E)=\frac{4}{52} \qquad P(F)=\frac{13}{52}$$

مثال: دو سکه را پرتاب می‌کنیم. فرض کنید ۴ نتیجه ممکن هم شانس باشند. اگر پیشامد E نشان دهنده اینکه اولین سکه شیر بیاید و پیشامد F این باشد که سکه دوم خط بیاید. آنگاه E و F مستقل‌اند: زیرا

$$P(E\cap F)=P\{HT\}=\frac{1}{4}$$

$$P(E)=P(HH,HT)=\frac{1}{2}$$

$$P(F)=P(HT,TT)=\frac{1}{2}$$

مثال: فرض کنید دو تاس منتظم را یک مرتبه پرتاب کنیم. اگر E_1 نشان دهنده پیشامد ۶ بودن جمع دو تاس و F نشان دهنده پیشامدی باشد که عدد تاس اول ۴ شود آنگاه

$$P(E_1\cap F)=P(\{(4,2)\})=\frac{1}{36}$$

$$P(E_1)P(F)=\frac{5}{36}\times\frac{1}{6}=\frac{5}{216}$$

بنابراین E_1 و F مستقل نیستند.

حال فرض کنید E_2 پیشامدی باشد که مجموع دو تاس برابر با ۷ است. آیا E_2 مستقل از F است؟

$$P(E_2 \cap F) = P(\{(4,3)\}) = \frac{1}{36}$$

$$P(E_2)P(F) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

مثال: اگر E نشان دهنده پیشامدی باشد که رئیس جمهور بعدی از حزب جمهوری‌خواه است و F نشان دهنده پیشامدی باشد که در طول سال آینده یک زلزله عظیم رخ خواهد داد. بسیاری از مردم مایل هستند که دو پیشامد E و F را مستقل از یکدیگر بدانند. در صورتی که اگر پیشامد G، وقوع یک جنگ بزرگ در دو سال بعد از انتخابات باشد، مسلماً مستقل فرض کردن دو پیشامد E و G مورد قبول همه مردم نخواهد بود

قضیه: اگر E و F مستقل باشند آنگاه E و F^c نیز مستقلند.

$$P(E \cap F^c) = P(E)P(F^c)$$

باید نشان دهیم

$$P(E \cap F^c) = P(E-F) = P(E) - P(E \cap F) \underset{F \text{ و } E \text{ با استقلال}}{=} P(E) - P(E)P(F)$$

$$= P(E)\left(1 - P(F)\right) = P(E)P(F^c)$$

مثال: دو تاس منتظم را پرتاب کرده‌ایم. اگر E نشان دهنده پیشامد اینکه مجموع دو تاس ۷ باشد و F پیشامدی باشد که تاس اول عدد ۴، و G پیشامد اینکه تاس دوم عدد ۳ است.

E: پیشامد اینکه مجموع دو تاس ۷ باشد $\qquad E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

F: پیشامد اینکه تاس اول عدد ۴ $\qquad P(E) = \frac{6}{36} = \frac{1}{6}$

G: پیشامد اینکه تاس دوم عدد ۳ $\qquad P(F) = \frac{6}{36} = \frac{1}{6}$

$$P(E \cap F) = P\{(4,3)\} = \frac{1}{36} = P(E)P(F) \qquad E \perp F$$

$$P(E \cap G) = P\{(4,3)\} = \frac{1}{36} = P(E)P(G) \qquad E \perp G$$

اما E مستقل از $F \cap G$ نیست زیرا

$$P(E \mid F \cap G) = 1 \neq P(E) = \frac{1}{6}$$

تعریف: سه پیشامد E، F و G مستقلند اگر

$$P(E \cap F \cap G) = P(E)P(F)P(G)$$

$$P(E \cap F) = P(E)P(F) \qquad P(E \cap G) = P(E)P(G) \qquad P(F \cap G) = P(F)P(G)$$

توضیح: اگر E، F و G مستقل باشند. آنگاه E مستقل از هر پیشامدی است که از F و G تشکیل شده باشند.

مثال: E مستقل از $F \cup G$ است، اگر E، F و G مستقل باشند

$$P(E \cap (F \cup G)) = P((E \cap F) \cup (E \cap G)) = P(E \cap F) + P(E \cap G)$$

$$- P(E \cap F \cap G) = P(E)P(F) + P(E)P(G) - P(E)P(G)P(F)$$

$$= P(E)\{P(F) + P(G) - P(F \cap G)\} = P(E)P(F \cup G)$$

تعریف استقلال را می‌توان به بیش از سه پیشامد تعمیم داد.

پیشامدهای E_1، E_2، ...، E_n مستقلند اگر برای هر زیرمجموعه $E_{1'}$، $E_{2'}$، ...، $E_{r'}$ $(r' \leq n)$ از پیشامدها داشته باشیم

$$P(E_{1'} \cap E_{2'} \cap \cdots E_{r'}) = P(E_{1'})P(E_{2'}) \cdots P(E_{r'}).$$

یک مجموعه نامتناهی از پیشامدها را مستقل گویند اگر هر زیرمجموعه متناهی از آنها مستقل باشد.

مثال: یک دنباله نامتناهی از آزمایشهای برنولی مستقل را انجام داده، در هر آزمایش برنولی نتیجه موفقیت را با احتمال p و نتیجه شکست را با احتمال $1-p$ در نظر می‌گیریم.

مطلوب است احتمال آنکه

الف- حداقل یک موفقیت در n آزمایش ساده اول کسب شود.

ب- دقیقاً k موفقیت در n آزمایش ساده اول به دست آید.

ج- همه آزمایشهای ساده (برنولی) نتیجه موفقیت داشته باشند.

E_i: پیشامد شکست در آزمایش برنولی iام

$$P(\text{عدم موفقیت}) = P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1)P(E_2)\cdots P(E_n)$$
$$= (1-p)^n$$

$$P(\text{حداقل یک موفقیت}) = 1-(1-p)^n$$

حل (ب) : هر دنباله مشخص از n آزمایش ساده اول را به صورت k موفقیت و $n-k$ شکست در نظر می‌گیریم. هر یک از دنباله‌ها دارای احتمال وقوع $p^k(1-p)^{n-k}$ است و چون $\binom{n}{k}$ از چنین دنباله‌هایی وجود دارد پس

$$P(\text{دقیقاً } k \text{ موفقیت}) = \binom{n}{k} p^k (1-p)^{n-k}$$

حل (ج)

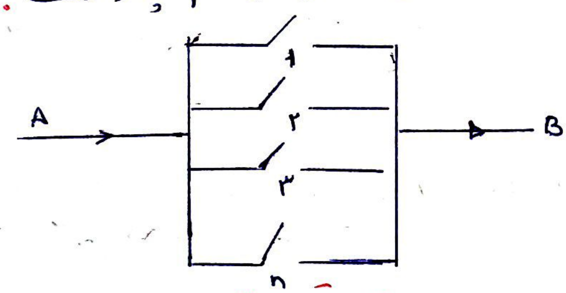
$$P(E_1^c \cap E_2^c \cap \dots \cap E_n^c) = p^n$$

با توجه به خاصیت پیوستگی تابع احتمال، احتمال مطلوب $P\left(\bigcap_{i=1}^{\infty} E_i^c\right)$ را به صورت زیر محاسبه می‌کنیم

$$P\left(\bigcap_{i=1}^{\infty} E_i^c\right) = P\left(\lim_{n\to\infty} \bigcap_{i=1}^{n} E_i^c\right)$$

$$= \lim_{n\to\infty} P\left(\bigcap_{i=1}^{n} E_i^c\right) = \lim_{n\to\infty} p^n = \begin{cases} 0 & p<1 \\ 1 & p=1 \end{cases}$$

مثال: سیستمی که از n جزء مجزا تشکیل شده است سیستم موازی نامیده می‌شود اگر برای کارکردن آن لازم باشد که حداقل یکی از اجزاء کار کند. برای چنین سیستمی اگر جزء iام مستقل از سایر اجزاء با احتمال p_i $(i=1,2,\dots,n)$ کار کند احتمال کارکرد سیستم چقدر است؟



سیستم موازی

A_i : پیشامد آنکه جزء iام کار کند

$$P(\text{سیستم کار کند}) = P\left(\bigcup_{i=1}^{n} A_i\right) = 1-P(\text{سیستم کار نکند}) = 1-P(\text{همه اجزاء سیستم کار نکنند})$$

$$= 1-P\left(\bigcap_{i=1}^{n} A_i^c\right) = 1-\prod_{i=1}^{n}(1-p_i)$$

مثال: آزمایش‌های ساده مستقل شامل پرتاب یک جفت تاس مستقل را در نظر می‌گیریم. احتمال اینکه نتیجه ۵ قبل از نتیجه ۷ برای مجموع دو تاس ظاهر شود را به دست آورید.

E_n: پیشامد ظاهر شدن مجموع ۵ و یا مجموع ۷ در $n-1$ آزمایش

ساده اول و ظاهر شدن مجموع ۵ در آزمایش ساده n ام

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

$$P\{\text{آمدن مجموع ۵ در هر آزمایش ساده}\} = P\{(1,4),(2,3),(3,2),(4,1)\} = \frac{4}{36}$$

$$P\{\text{آمدن مجموع ۷ در هر آزمایش ساده}\} = P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \frac{6}{36}$$

$$P(E_n) = \left(1-\frac{10}{36}\right)^{n-1}\frac{4}{36}$$

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty}\left(1-\frac{10}{36}\right)^{n-1}\frac{4}{36} = \frac{1}{9}\sum_{n=1}^{\infty}\left(\frac{13}{18}\right)^{n-1} = \frac{1}{9}\times\frac{1}{1-\frac{13}{18}} = \frac{2}{5}$$

می‌توانستیم این نتیجه را با استفاده از احتمال شرطی نیز به دست آوریم.

E: پیشامد آمدن مجموع ۵ پیش از مجموع ۷

$P(E)$ را می‌توان با شرطی کردن روی نتیجه اولین آزمایش ساده به دست آورد.

$P(E) = ?$

F: پیشامد آنکه نتیجه اولین آزمایش ساده برابر ۵ باشد

G: 〃 〃 〃 〃 〃 برابر ۷ باشد

H: پیشامد آنکه نتیجه اولین آزمایش ۵ و ۷ نباشد

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

$$\Rightarrow P(E) = 1\times\frac{4}{36} + 0\times\frac{6}{36} + P(E)\times\frac{26}{36}$$

$$\Rightarrow P(E) = \frac{1}{9} + \frac{13}{18}P(E) \Rightarrow \frac{5}{18}P(E) = \frac{1}{9} \Rightarrow P(E) = \frac{2}{5}$$

خواننده باید توجه داشته باشد که پاسخ کاملاً شهودی است. چون آمدن ۵ در هر آزمایش ساده دارای احتمال $\frac{4}{36}$ و آمدن ۷ دارای احتمال $\frac{6}{36}$ است، به نظر می‌رسد که شانس آمدن ۵ پیش از ۷ باید به نسبت ۴ به ۶ باشد، یعنی احتمال برابر با $\frac{4}{10}$ است.

همین استدلال نشان می‌دهد که اگر E و F دو پیشامد ناسازگار باشند آنگاه وقتی که آزمایش‌های ساده مستقل از چنین آزمایشی انجام می‌گیرد پیشامد E با احتمال

$$\frac{P(E)}{P(E)+P(F)}$$

پیش از F رخ می‌دهد.

مسأله نابودی قمارباز:

دو قمارباز A و B روی نتیجه پرتاب یک سکه شرط‌بندی می‌کنند. در هر پرتاب اگر سکه شیر بیاید A یک تومان از B می‌گیرد و اگر خط ظاهر شود A یک تومان به B می‌دهد. آنها آنقدر بازی را ادامه می‌دهند تا یکی از آنها پولش تمام شود اگر فرض شود که پرتاب‌های متوالی مستقل هستند و احتمال شیر آمدن p باشد احتمال اینکه A همه پول‌ها را ببرد، به شرط اینکه وی با i و B با $N-i$ تومان شروع کرده باشند را به دست آورید.

حل:

E: نشان دهنده آن که A بازی را ببرد درصورتی که از آغاز A i تومان و B $N-i$ تومان داشته باشد

قرار دهید $P_i = P(E)$ بدین دلیل است که احتمال برد بازیکن A به سرمایه اولیه که i تومان است بستگی دارد

حال برای محاسبه $P(E)$ روی اولین پرتاب شرط می‌کنیم. اگر H پیشامد آن باشد که نتیجه پرتاب نخست شیر است آنگاه

$$P_i = P(E) = P(E|H)\,P(H) + P(E|H^c)\,P(H^c)$$

$$= p\,P(E|H) + (1-p)\,P(E|H^c)$$

اگر بدانیم که اولین پرتاب شیر است، وضعیت بعد از اولین پرتاب این است که A مبلغ $i+1$

تومان و B مبلغ $N-(i+1)$ تومان پول دارند. چون پرتاب‌ها مستقل اند و احتمال شیر آمدن در هر پرتاب برابر با p است، بنابراین از این نقطه به بعد احتمال بردن همه پول توسط A دقیقاً برابر است با احتمال بردن با این وضعیت که بازی تازه شروع شده و A سرمایه‌اش $i+1$ و B سرمایه‌اش $N-(i+1)$ تومان است. بنابراین $P(E|H) = P_{i+1}$ و $P(E|H^c) = P_{i-1}$.

$$\Rightarrow P_i = p\,P_{i+1} + (1-p)\,P_{i-1} = p\,P_{i+1} + q\,P_{i-1} \qquad i = 1, 2, \ldots, N-1$$

حال می‌خواهیم معادله $P_i = p\,P_{i+1} + q\,P_{i-1}$ را با شروط اولیه $P_0 = 0$ و $P_N = 1$ و $p+q=1$ حل کنیم. بنابراین

$$(p+q)\,P_i = p\,P_{i+1} + q\,P_{i-1} \Rightarrow P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}) \qquad i = 1, \ldots, N-1$$

بنابراین با قرار دادن $i=1$ در رابطه فوق داریم

$$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) \underset{P_0=0}{=} \frac{q}{p} P_1$$

به همین ترتیب به ازای $i=2$، $i=3$، ... به دست می‌آوریم

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

$$\vdots$$

$$P_i - P_{i-1} = \frac{q}{p}(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1$$

$$\vdots$$

$$P_N - P_{N-1} = \frac{q}{p}(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1$$

با جمع i-1 معادله اول داریم

$$P_i - P_1 = P_1\left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right]$$

حال اگر $\frac{q}{p} = 1$ آنگاه

$$P_i - P_1 = P_1(i-1) \Rightarrow P_i = i\,P_1$$

و اگر $\frac{q}{p} \neq 1$، آنگاه

$$P_i - P_1 = P_1 \times \frac{\frac{q}{p} - \left(\frac{q}{p}\right)\left(\frac{q}{p}\right)^{i-1}}{1-\frac{q}{p}} \Rightarrow P_i = P_1\left(1 + \frac{\frac{q}{p} - \left(\frac{q}{p}\right)^i}{1-\frac{q}{p}}\right)$$

$$= P_1\left(\frac{1-\frac{q}{p}+\frac{q}{p}-\left(\frac{q}{p}\right)^i}{1-q/p}\right) = P_1 \times \frac{1-\left(\frac{q}{p}\right)^i}{1-q/p}$$

$$P_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)} P_1 & \dfrac{q}{p} \neq 1 \\ i P_1 & \dfrac{q}{p} = 1 \end{cases} \qquad (*)$$

از طرف دیگر با جمع $N-1$ معادله اول می‌توان به دست آورد

$$P_N - P_1 = P_1 \left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{N-1} \right]$$

اما چون $P_N = 1$ لذا برای $\frac{q}{p} \neq 1$ و یا برای $\frac{1-p}{p} \neq 1$ که معادل با $p \neq \frac{1}{2}$ داریم

$$1 - P_1 = P_1 \times \frac{\frac{q}{p} - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} \Rightarrow P_1 \times \left(\frac{\frac{q}{p} - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} + 1 \right) = 1$$

$$\Rightarrow P_1 \times \frac{1 - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} = 1 \Rightarrow P_1 = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}$$

همچنین در صورتی که $p = q$ و یا $p = \frac{1}{2}$ باشد آنگاه $P_N = N P_1$ و یا $P_1 = \frac{1}{N}$ است. پس به طور خلاصه

$$P_1 = \begin{cases} \dfrac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N} & p \neq \frac{1}{2} \\ \dfrac{1}{N} & p = \frac{1}{2} \end{cases}$$

و بنابراین با جایگزین کردن P_1 در رابطه $(*)$ به دست می‌آوریم

$$P_i = \begin{cases} \dfrac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N} & p \neq \frac{1}{2} \\ \dfrac{i}{N} & p = \frac{1}{2} \end{cases}$$

حال فرض کنید که Q_i نشان دهنده احتمال این باشد که B همه پولها را ببرد وقتی که A با i تومان و B با $N-i$ تومان شروع کرده باشد، آنگاه با استفاده از تقارن و جایگزین کردن p با q و i با $N-i$ داریم

$$Q_i = \begin{cases} \dfrac{1 - \left(\frac{p}{q}\right)^{N-i}}{1 - \left(\frac{p}{q}\right)^N} & q \neq \frac{1}{2} \\ \dfrac{N-i}{N} & q = \frac{1}{2} \end{cases}$$

حال نشان می‌دهیم که $P_i+Q_i=1$ زیرا در حالتی که $p=q=\frac{1}{2}$ داریم $P_i+Q_i=\frac{i}{N}+\frac{N-i}{N}=1$ و هنگامی که $p\neq\frac{1}{2}$ و به طبع آن $q\neq\frac{1}{2}$ خواهیم داشت

$$P_i+Q_i=\frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N}+\frac{1-\left(\frac{p}{q}\right)^{N-i}}{1-\left(\frac{p}{q}\right)^N}=\frac{p^N-(q/p)^i p^N}{p^N-q^N}+\frac{q^N-q^N\left(\frac{p}{q}\right)^{N-i}}{q^N-p^N}$$

$$=\frac{p^N-\cancel{p^{N-i}q^i}-q^N+\cancel{q^i p^{N-i}}}{p^N-q^N}=\frac{p^N-q^N}{p^N-q^N}=1$$

از این جهت $P_i+Q_i=1$، لذا با احتمال یک، یکی از دو بازیکن A یا B بالاخره باید همه پولها برنده می‌شود یا به عبارتی اگر بازی به طور نامحدود ادامه یابد احتمال اینکه دارایی A بین ۱ و $N-1$ باشد صفر است. باید توجه شود که نتیجه این آزمایش سه حالت است. A برنده شود یا B برنده شود و یا اینکه بازی برنده نداشته باشد که نشان دادیم احتمال نتیجه سوم، صفر است.

به عنوان مثال اگر A با ۵ تومان و B با ۱۰ تومان شروع کنند و $p=\frac{1}{2}$ باشد آنگاه احتمال برد A برابر $\frac{1}{3}$ است و این احتمال برای $p=0.6$ برابر است با

$$\frac{1-\left(\frac{0.4}{0.6}\right)^5}{1-\left(\frac{0.4}{0.6}\right)^{5+10}}=\frac{1-\left(\frac{2}{3}\right)^5}{1-\left(\frac{2}{3}\right)^{15}}=0.87$$

یک حالت خاص مسأله نابودی قمارباز که به عنوان مسأله «مدت بازی» معروف است در سال ۱۶۵۷ توسط فرما به ریاضیدان آلمانی، کریستین هویگنس، داده شد. مسأله این بود که A و B هرکدام ۱۲ سکه دارند. آنها برای این سکه‌ها با ۳ تاس یک بازی را به صورت زیر شروع می‌کنند. هر وقت مجموع اعداد مشاهده شده برابر ۱۱ باشد (فرقی نمی‌کند که چه کسی تاسها را پرتاب کند) آنگاه A یک سکه به B می‌دهد، هر وقت عدد ۱۴ مشاهده شد B یک سکه به A می‌دهد. شخصی که اول همه سکه‌ها را برد برنده بازی است.

چون $P(\text{نتیجه } 11) = \frac{27}{216}$ و $P(\text{نتیجه } 14) = \frac{15}{216}$ ، بنابراین احتمال آنکه مجموع ۱۴ پیش از مجموع ۱۱ مشاهده شود $p = \frac{15}{15+27}$ یعنی $p = \frac{15}{42}$ است. در نتیجه با توجه به مسئله نابودی قمارباز داریم $p = \frac{15}{42}$ ، $i = 12$ ، $N = 24$. حالت عمومی مسئله نابودی قمارباز توسط جیمز برنولی ریاضیدان معروف حل شد و ۸ سال پس از وفاتش در سال ۱۷۱۳ منتشر گردید.

مثال: یک گراف کامل که دارای n رأس باشد تشکیل شده است از n نقطه (رأس) در صفحه و $\binom{n}{2}$ خط (یال) که هر زوج رأسها را به هم متصل نموده است. یک گراف کامل با ۳ رأس در شکل زیر نشان داده شده است. فرض کنید هر یال در یک گراف کامل n رأسی را به رنگ قرمز یا آبی نشان دهیم. برای یک عدد ثابت K ، سؤال جالب این است که آیا راهی برای رنگ کردن یالها وجود دارد که هیچ مجموعه K رأسی همه $\binom{K}{2}$ یالش هم رنگ نباشند؟ با یک استدلال احتمالی می‌توان نشان داد که اگر n چندان بزرگ نباشد آنگاه پاسخ مثبت است.

استدلال چنین است که اگر هر یال مستقل از دیگر یالها با شانس برابر رنگ قرمز یا رنگ آبی داشته باشد، یعنی هر یال با احتمال $\frac{1}{2}$ قرمز باشد، تعداد $\binom{n}{K}$ مجموعه‌های K رأسی را شمارش نموده و پیشامدهای E_i $\left(i = 1, 2, \dots, \binom{n}{K}\right)$ را به صورت زیر تعریف می‌کنیم:

$E_i = \{$همه یالهای وصل کننده i امین مجموعه K رأسی هم‌رنگ باشند$\}$

حال چون هر یک از $\binom{K}{2}$ یال وصل کننده مجموعه K رأسی با شانس برابر قرمز یا آبی است
در نتیجه احتمال اینکه همه آنها از یک رنگ باشند برابر است با

$$P(E_i) = 2\left(\frac{1}{2}\right)^{\frac{K(K-1)}{2}}$$

حال $P\left(\bigcup_i E_i\right)$ یعنی احتمال اینکه حداقل یک مجموعه K رأسی همه یالهایش هم رنگ باشند
با استفاده از نامساوی بول، در نامساوی زیر صدق می‌کند:

$$P\left(\bigcup_i E_i\right) \leq \sum_i P(E_i) = \binom{n}{k}\left(\frac{1}{2}\right)^{\frac{k(k-1)}{2}-1}$$

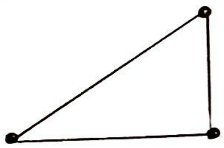
زیرا اولاً $P(E_i)$ به اندیس i بستگی ندارد و نیز تعداد جملات $P(E_i)$ ها، تعداد مجموعه k رأسی از میان n رأس یعنی $\binom{n}{k}$ است.

بنابراین اگر

$$\binom{n}{k}\left(\frac{1}{2}\right)^{\frac{k(k-1)}{2}-1} \iff \binom{n}{k} < 2^{\frac{k(k-1)}{2}-1}$$

آنگاه احتمال اینکه حداقل یکی از $\binom{n}{k}$ مجموعه k رأسی دارای یالهای هم‌رنگ باشد کمتر از ۱ است.

بنابراین با شرایط فوق n و k نتیجه می‌شود که با احتمال مثبت هیچ مجموعه k رأسی وجود ندارد که یالهایش از یک رنگ باشند، و این موضوع نتیجه می‌دهد حداقل یک راه برای اینکه هیچ مجموعه k رأسی همه یالهایش هم‌رنگ نباشد وجود دارد.



تذکر: باید توجه داشت که استدلال فوق اگرچه روی n و k شرط می‌گذارد تا وجود یک طرح را که دارای خصوصیات لازم است بیان دارد، اما هیچ اطلاعی درباره اینکه چگونه چنین طرحی تهیه شود ارائه نمی‌دهد. (اگرچه چنین امکانی می‌تواند به سادگی در انتخاب تصادفی رنگها، کنترل نتیجه رنگها، کنترل نتیجه رنگها و تکرار آن تا حصول نتیجه ثبت کند)

$P(\cdot|F)$ به عنوان یک تابع احتمال

احتمالهای شرطی همه خصوصیات یک احتمال معمولی را دارا می‌باشند.

گزاره:

الف) $0 \leq P(E|F) \leq 1$

ب) $P(S|F) = 1$

ج) اگر E_i ها $(i=1,2,\ldots)$ پیشامدهای دو به دو ناسازگار باشند آنگاه

$$P\left(\bigcup_{i=1}^{\infty} E_i \,\middle|\, F\right) = \sum_{i=1}^{\infty} P(E_i|F)$$

اثبات: برای اثبات قسمت (الف) باید نشان دهیم که $0 \le \frac{P(E\cap F)}{P(F)} \le 1$ که طرف چپ نامساوی واضح است زیرا $P(F)\neq 0$ و $P(E\cap F)\geq 0$ است. برای کران بالا چون $E\cap F \subseteq F$ لذا داریم $P(E\cap F) \le P(F)$ و از این رو $\frac{P(E\cap F)}{P(F)} \le 1$.

برای قسمت ب:

$$P(S|F) = \frac{P(S\cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

اثبات قسمت ج:

$$P\left(\bigcup_{i=1}^{\infty} E_i \,\Big|\, F\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} E_i\right)\cap F\right)}{P(F)} = \frac{P\left(\bigcup_{i=1}^{\infty} (E_i\cap F)\right)}{P(F)}$$

$$= \frac{\sum_{i=1}^{\infty} P(E_i\cap F)}{P(F)} = \sum_{i=1}^{\infty} \frac{P(E_i\cap F)}{P(F)} = \sum_{i=1}^{\infty} P(E_i|F)$$

زیرا $E_i\cap F$ ها دو به دو ناسازگار هستند

$$\forall i\neq j \quad (E_i\cap F)\cap(E_j\cap F) = E_i\cap E_j\cap F = \emptyset$$

اگر $P(E|F)=Q(E)$ آنگاه با توجه به گزاره فوق می‌توان $Q(E)$ را به عنوان یک تابع احتمال روی پیشامدهای S در نظر گرفت و بنابراین همه قضایایی که برای قضایای احتمال برقرار است، برای قضایای احتمال شرطی نیز برقرار است.

$$Q_1(E_1\cup E_2) = Q(E_1) + Q(E_2) - Q(E_1\cap E_2)$$

یا به طور معادل

$$P(E_1\cup E_2|F) = P(E_1|F) + P(E_2|F) - P(E_1\cap E_2|F).$$

همچنین اگر احتمال شرطی را به صورت $Q(E_1|E_2)$ تعریف کنیم آنگاه

$$Q(E_1|E_2) = \frac{Q(E_1\cap E_2)}{Q(E_2)}$$

همچنین

$$Q(E_1) = Q(E_1|E_2)Q(E_2) + Q(E_1|E_2^c)Q(E_2^c) \quad (**)$$

$$Q(E_1|E_2) = \frac{Q(E_1\cap E_2)}{Q(E_2)} = \frac{P(E_1\cap E_2|F)}{P(E_2|F)} = \frac{\frac{P(E_1\cap E_2\cap F)}{P(F)}}{\frac{P(E_2\cap F)}{P(F)}}$$

$$= \frac{P(E_1\cap E_2\cap F)}{P(E_2\cap F)} = P(E_1|E_2\cap F)$$

درحقیقت رابطهٔ (**) معادل رابطهٔ زیر است

$$P(E_1|F) = P(E_1|E_2 \cap F)\,P(E_2|F) + P(E_1|E_2^c \cap F)\,P(E_2^c|F)$$

مثال: شرکت بیمه ای را در نظر بگیرید که اعتقاد دارد، مردم را می توان به دو دستهٔ مجزا، آنهایی که مستعد تصادف هستند و آنهایی که نیستند تقسیم نمود. در هر سال معین یک فرد مستعد تصادف با احتمال ۰٫۴ تصادف خواهد داشت درحالی که این عدد برای فرد غیر مستعد برابر با ۰٫۲ است. احتمال شرطی اینکه یک فرد بیمه شده در دومین سال قرارداد تصادف داشته باشد به شرط اینکه در سال اول تصادف داشته است را به دست آورید.

A: پیشامد اینکه فرد بیمه شده مستعد تصادف باشد

A_1: پیشامد آنکه بیمه شده در اولین سال یک تصادف داشته باشد

A_2: پیشامد آنکه بیمه شده در دومین سال یک تصادف داشته باشد

$$P(A_2|A_1) = ?$$

$$P(A_2|A_1) = P(A_2|A_1 \cap A)\,P(A|A_1) + P(A_2|A_1 \cap A^c)\,P(A^c|A_1)$$

حال آنکه

$$P(A|A_1) = \frac{P(A_1|A)\,P(A)}{P(A_1)} = \frac{(0/4)(0/3)}{0/26} = \frac{6}{13}$$

در مثال قبلاً با استفاده از قضیه احتمال کل نشان دادیم که $P(A_1) = 0/26$

$$P(A^c|A_1) = 1 - P(A|A_1) = 1 - \frac{6}{13} = \frac{7}{13}$$

همچنین $P(A_2|A_1 \cap A) = 0/4$ و $P(A_2|A^c \cap A_1) = 0/2$، بنابراین

$$P(A_2|A_1) = (0/4)\frac{6}{13} + (0/2)\frac{7}{13} \approx 0/29$$

در اینجا پیشامدهای اینکه فرد بیمه شده در سال ($i=1,2$) یک حادثه داشته باشد را به شرط پیشامد اینکه فرد مورد نظر مستعد حادثه یا غیر مستعد باشد مستقل در نظر گرفته ایم.

مثال: آزمایش های ساده مستقل هر کدام با احتمال موفقیت p و احتمال شکست $q = 1-p$ انجام می گیرند. می خواهیم احتمال وقوع یک دنباله n تایی از موفقیتهای متوالی را قبل از m شکست پیاپی به دست آوریم.

حل: از روی کتاب مطالعه شود.

مثال: در یک مهمانی n مرد کلاه‌های خود را درآورده و پس از مخلوط شدن آنها هر کدام به تصادف کلاهی را انتخاب می‌کنند. هرگاه مردی کلاه خود را انتخاب کند گوییم یک انطباق اتفاق افتاده است.

الف- احتمال اینکه هیچ انطباق اتفاق نیفتد چقدر است؟

ب- احتمال اینکه دقیقاً k انطباق اتفاق افتد چقدر است؟

حل: الف) فرض کنید E نشان دهنده پیشامدی باشد که هیچ انطباق اتفاق نیفتد، آنگاه

$$P(E) = P_n$$

حال با مشروط کردن اینکه اولین مرد کلاه خود را انتخاب می‌کند یا خیر، شروع می‌کنیم

M : پیشامد آنکه فرد نخست کلاه خود را انتخاب می‌کند

$$P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c)$$

واضح است $P(E|M) = 0$ زیرا فرد اول کلاه خودش را برداشته است.

$$P_n = P(E|M^c) \times \frac{n-1}{n} \qquad (***)$$

اما $P(E|M^c)$ احتمال پیشامدی است که برای $n-1$ مرد باقی مانده از $n-1$ کلاه باقی مانده که کلاه یکی از این مردها در بین آنها نیست، کلاهی را به تصادف انتخاب کنند و هیچ انطباق اتفاق نیفتد. این پیشامد به دو حالت مجزا اتفاق می‌افتد، هیچ انطباقی نیست و مرد اضافی کلاه اضافی را انتخاب نمی‌کند (این کلاه مردی است که اول کلاهش را انتخاب کرده است) یا اینکه مرد اضافی کلاه اضافی را انتخاب می‌کند. احتمال اولین حالت همان P_{n-1} است که به صورت تصور اینکه کلاه اضافی کلاه مرد اضافی است، دیده می‌شود. دومین پیشامد دارای احتمال $\frac{1}{n-1}P_{n-2}$ است و داریم

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1}P_{n-2}$$

$$\overset{(***)}{\Rightarrow} \quad P_n = \frac{n-1}{n}P_{n-1} + \frac{1}{n}P_{n-2} \Rightarrow P_n - P_{n-1} = \frac{-1}{n}(P_{n-1} - P_{n-2})$$

چون P_n احتمال این است که هرگاه n مرد n کلاه را انتخاب می‌کنند، هیچ انطباق اتفاق نیفتد پس

$$P_1 = 0 \qquad P_2 = \frac{1}{2}$$

پس

$$P_3 - P_2 = -\frac{P_2 - P_1}{3} = -\frac{1}{3!} \qquad P_3 = \frac{1}{2!} - \frac{1}{3!}$$

$$P_4 - P_3 = -\frac{P_3 - P_2}{4} = \frac{1}{4!} \qquad P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

پس در حالت کلی

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}$$

ب - برای به دست آوردن احتمال k انطباق اگر گروه ثابت k تایی از مردان را در نظر می‌گیریم، احتمال اینکه فقط آنها کلاه‌های خود را انتخاب کنند برابر است با

$$\frac{1}{n} \times \frac{1}{n-1} \times \cdots \times \frac{1}{n-(k-1)} P_{n-k} = \frac{(n-k)!}{n!} P_{n-k}$$

که P_{n-k} احتمال شرطی این است که برای n-k مرد دیگر که کلاه انتخاب می‌کنند هیچ انطباق وجود نداشته باشد. چون k مرد را به $\binom{n}{k}$ طریق مختلف می‌توان انتخاب نمود پس احتمال اینکه دقیقاً k انطباق اتفاق افتد برابر است با

$$\frac{P_{n-k}}{k!} = \frac{\frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{n-k}}{(n-k)!}}{k!}$$

نکته: استقلال پیشامدهای شرطی

گوییم پیشامدهای E_1 و E_2 به شرط پیشامد F، مستقل هستند اگر در صورت وقوع پیشامد F، احتمال شرطی اینکه E_1 اتفاق افتد با اطلاع از وقوع یا عدم وقوع E_2 تغییر نیابد. به عبارتی E_1 و E_2 به شرط F مستقل هستند هرگاه

$$P(E_1 | E_2 \cap F) = P(E_1 | F)$$

یا به طور معادل

$$P(E_1 \cap E_2 | F) = P(E_1 | F)\, P(E_2 | F)$$

مفهوم فوق را می‌توان به بیش از دو پیشامد نیز تعمیم داد.

مثال: قاعده توالی لاپلاس فرض کنید $(k+1)$ سکه در یک جعبه قرار دارند، اگر i امین سکه را پرتاب کنیم با احتمال i/k $(i=0,1,\dots,k)$ شیر می‌آید. یک سکه را به تصادف از جعبه انتخاب نموده و آن را مکرراً پرتاب می‌کنیم. اگر n پرتاب اول نتیجه‌اش شیر باشد احتمال شرطی اینکه $(n+1)$ امین پرتاب نیز شیر باشد را به دست آورید.

حل: فرض کنید:

C_i : نشان دهنده این باشد که i امین سکه به تصادف انتخاب شده، $i=0,1,\dots,k$

H : پیشامد آنکه $(n+1)$ امین پرتاب نیز شیر باشد

F_n : پیشامد آنکه n پرتاب اول همگی شیر باشند

$$P(H\mid F_n)=?$$

$$P(H\mid F_n)=\sum_{i=0}^{k} P(H\mid F_n\cap C_i)\,P(C_i\mid F_n)$$

حال اگر بدانیم که سکه iام انتخاب شده، شفاف است که فرض کنیم نتایج به صورت مشروط مستقل هستند و هر نتیجه با احتمال i/k شیر خواهد بود بنابراین

$$P(H\mid F_n\cap C_i)=P(H\mid C_i)=\frac{i}{k}$$

$$P(C_i\mid F_n)=\frac{P(C_i\cap F_n)}{P(F_n)}=\frac{P(F_n\mid C_i)\,P(C_i)}{\sum_{j=0}^{k}P(F_n\mid C_j)\,P(C_j)}$$

$$=\frac{\left(\frac{i}{k}\right)^n\left[\frac{1}{k+1}\right]}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^n\left[\frac{1}{k+1}\right]}$$

بنابراین

$$P(H\mid F_n)=\frac{\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1}}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^{n}}$$

اما اگر k بزرگ باشد با استفاده از تقریب انتگرال داریم

$$\frac{1}{k}\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1}\approx\int_0^1 x^{n+1}\,dx=\frac{1}{n+2}$$

$$\frac{1}{k}\sum_{j=0}^{k}\left(\frac{j}{k}\right)^n \approx \int_0^1 x^n dx = \frac{1}{n+1}$$

و بنابراین برای مقادیر بزرگ k داریم

$$P(H \mid F_n) \approx \frac{n+1}{n+2}$$