Small examples in Astronomy

1. Two QSOs at different redshift are beside each other on the sky. Remarkable! Calculate probability: it is conditional on having noticed this at the start. Thus prob(A|A) = 1, consistent with our measure of belief in something we know.

2. Now calculate probability of finding a galaxy and a quasar within *r* of each other. We search the solid angle Ω and have already found surface densities ζ_G and ζ_Q. On finding a galaxy, we search around it for a quasar. We need

$$\mathrm{prob}(\mathrm{G\ in\ field\ and\ Q\ within}\ r) = \mathrm{prob}(\mathrm{Q\ within}\ r \mid \mathrm{G\ in\ field})\mathrm{prob}(\mathrm{G\ in\ field})$$

Assumes probabilities are independent – and this is what we want to test. Without resorting to models:

$$\mathrm{prob}(\mathrm{G\ in\ field}) = \varsigma_G \Omega \qquad \mathrm{prob}(\mathrm{Q\ within}\ r) = \pi r^2 \varsigma_Q.$$

Thus

$$\mathrm{prob}(\mathrm{G\ in\ field\ and\ Q\ within}\ r) = \varsigma_G \varsigma_Q \Omega \pi r^2.$$

…..symmetrical in quasar and galaxy surface densities: we could search first for a galaxy or for a quasar. Note strong dependence on search area – specify this before the experiment!

Small examples in Astronomy

We want to know, given the data, what is the probability/belief state of our model. Priors can change anticipated results in violent and dramatic ways.

Before 1987, 4 naked-eye SNe had been recorded in ten centuries. What, before 1987, was the probability ρ of a bright SN happening in the 20th century?

God's-eye viewpoint: Meaningless! Events are either certain or forbidden.
God does not play dice…

Data: 4 SNe in 10 centuries. Prior on ρ: total ignorance, uniform on [0, 1].
Model: Binomial, in any century we either get a SN or not (neglecting here the possibility of 2 or more SNe). Posterior probability is then

$$\mathrm{prob}(\rho \mid \mathrm{data}) \propto \binom{10}{4} \rho^4 (1-\rho)^6 \times \text{prior on } \rho.$$

Normalize it,

$$\int_0^1 \mathrm{prob}(\rho \mid \mathrm{data})\, \mathrm{d}\rho = 1, \qquad \int_0^1 \binom{10}{4} \rho^4 (1-\rho)^6\, \mathrm{d}\rho,$$

Using Gamma and Beta function,

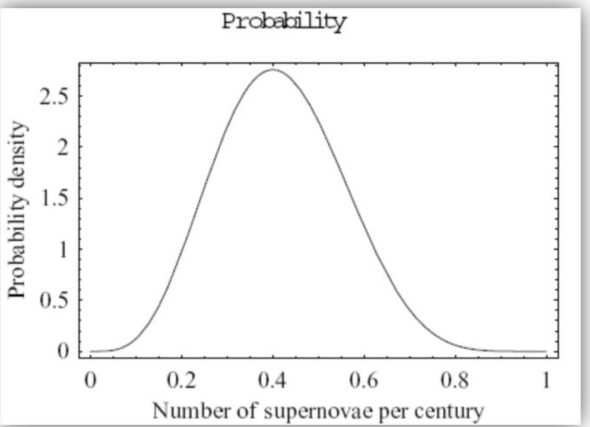
$$\frac{\Gamma(10)\Gamma(4)}{\Gamma(14)} = B[5, 7],$$

Small examples in Astronomy

In general, for *n* SNe in *m* centuries, posterior prob. Is

$$\text{prob}(\rho \mid \text{data}) = \frac{\rho^n (1-\rho)^{m-n}}{B[n+1, m-n+1]}.$$

Here n=4, m=10. Where does it peak?



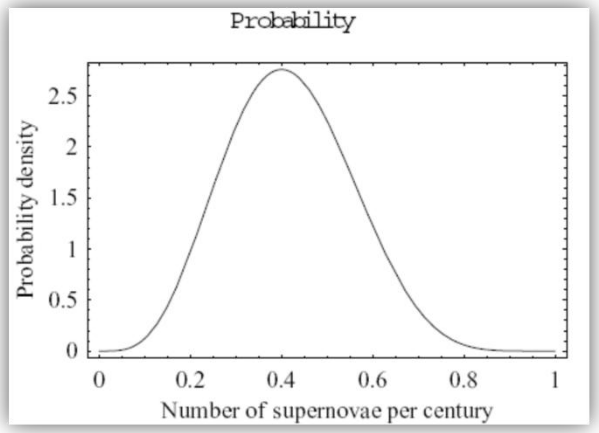
But… That was before 1987, say, at the end of 20th century… What about now?

Small examples in Astronomy

In general, for n SNe in m centuries, posterior prob. Is

$$\mathrm{prob}(\rho \mid \mathrm{data}) = \frac{\rho^n (1-\rho)^{m-n}}{B[n+1, m-n+1]}.$$

Here n=4, m=10. Where does it peak?



But… That was before 1987, say, at the end of 20th century… What about now?

Prior: $$\rho^4 (1-\rho)^6 / B[5, 7]$$

New data: exactly one event of prob. ρ.

Updated posterior:

$$\mathrm{prob}(\rho \mid \mathrm{data}) = \frac{\rho^5 (1-\rho)^6}{B[6, 7]}$$

Small examples in Astronomy

For a long time, objection to Bayesian focused on Bayes/Laplace uniform prior.

Jeffreys (1961), Jaynes (1968): in many cases that's far too agnostic. Intricate arguments led them to other possibilities:

$$\mathrm{prob}(\rho) = \frac{1}{\rho(1-\rho)}$$

$$\mathrm{prob}(\rho) = \frac{1}{\sqrt{\rho(1-\rho)}}.$$

'Haldane prior'

Small examples in Astronomy

For a long time, objection to Bayesian focused on Bayes/Laplace uniform prior.

Jeffreys (1961), Jaynes (1968): in many cases that's far too agnostic. Intricate arguments led them to other possibilities:

$$\mathrm{prob}(\rho) = \frac{1}{\rho(1-\rho)}$$

$$\mathrm{prob}(\rho) = \frac{1}{\sqrt{\rho(1-\rho)}}. \quad \text{'Haldane prior'}$$

We have focused on the peak – many other ways to characterize by a single number.
Posterior mean:

$$< \rho > = \int_0^1 \rho \, \mathrm{prob}(\rho \mid \mathrm{data}) \, \mathrm{d}\rho.$$

Small examples in Astronomy

For a long time, objection to Bayesian focused on Bayes/Laplace uniform prior.

Jeffreys (1961), Jaynes (1968): in many cases that's far too agnostic. Intricate arguments led them to other possibilities:

$$\mathrm{prob}(\rho) = \frac{1}{\rho(1-\rho)}$$

$$\mathrm{prob}(\rho) = \frac{1}{\sqrt{\rho(1-\rho)}}. \quad \text{'Haldane prior'}$$

We have focused on the peak – many other ways to characterize by a single number. **Posterior mean**:

$$<\rho> = \int_0^1 \rho\,\mathrm{prob}(\rho\,|\,\mathrm{data})\,\mathrm{d}\rho.$$

If we have had N successes and M failures, the posterior mean is given by the famous **Laplace's rule of succession**:

$$\boxed{<\rho> = \frac{N+1}{N+M+2}.}$$

e.g. M=0: "The Sun Also Rises"
e.g. SNe: In the year of 1899, N=4, N+M=10, predicts 5/12 rather than 4/10 (peak).

Characterization by a single number can be often misleading… unless the posterior distribution is very narrow (large samples).

Functions of jointly distributed random variables

PROPOSITION **B**

Let X be a continuous random variable with density $f(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I. Suppose that $f(x) = 0$ if x is not in I. Then Y has the density function

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

for y such that $y = g(x)$ for some x, and $f_Y(y) = 0$ if $y \neq g(x)$ for any x in I. Here g^{-1} is the inverse function of g; that is, $g^{-1}(y) = x$ if $y = g(x)$. ■

Special case A: sums

X, *Y* are discrete random variables, taking integer values and having joint frequency function $p(x, y)$. What is the frequency function of $Z=X+Y$?

Whenever $X=x$ and $Y=z-x$, $Z=z$, so the probability that $Z=z$ is the sum over all x of these joint probabilities,

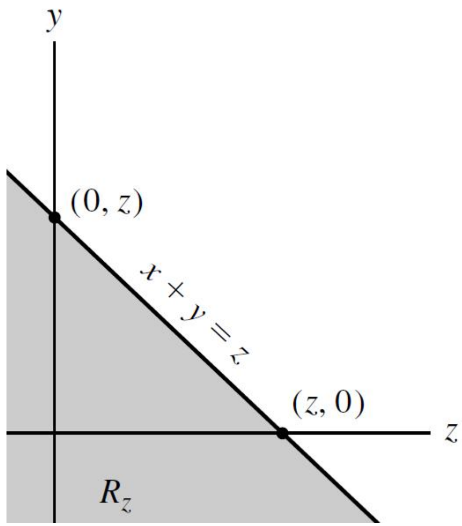
$$p_Z(z) = \sum_{x=-\infty}^{\infty} p(x, z-x)$$

If *X* and *Y* are independent,

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_X(x) p_Y(z-x)$$

- **convolution** of the sequences p_X and p_Y.

Special case A: sums



Continuous case

Find cdf of Z first!

$$F_Z(z) = \iint_{R_z} f(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y)\, dy\, dx$$

Differentiating it, using the rule of chains,

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\, dx$$

If X and Y are independent,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

- convolution of the functions f_X and f_Y.

Convolution is everywhere in astronomy:

- Smooth an image with a PSF
- Deriving the star formation history
- combination of multiple effects
- …

Special case A: sums

Continuous case

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

Example: The lifetime of a component is exponentially distributed. We have an identical and independent backup component.

Lifetime of the system $S=T_1+T_2$, its pdf is

$$f_S(s) = \int_0^{s} \lambda e^{-\lambda t} \lambda e^{-\lambda(s-t)} dt$$

Beyond the limits of integration, both components have 0 density.

$$f_S(s) = \lambda^2 \int_0^s e^{-\lambda s}\, dt$$
$$= \lambda^2 s e^{-\lambda s}$$

$$g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \qquad t \geq 0$$

Gamma distribution with parameters 2, λ.

Special case B: quotients

Find cdf of $Z=Y/X$ first, again.

$F_Z(z) = P(Z \leq z)$ is the probability of the set of (x, y) such that $y/x \leq z$. If $x > 0$, it is the set $y \leq xz$; if $x < 0$, it is the set $y \geq xz$.

$$F_Z(z) = \int_{-\infty}^{0} \int_{xz}^{\infty} f(x, y)\, dy\, dx + \int_{0}^{\infty} \int_{-\infty}^{xz} f(x, y)\, dy\, dx$$

Differentiating it, using the rule of chains, insert $y=xv$ (instead of xz),

$$\begin{aligned}
F_Z(z) &= \int_{-\infty}^{0} \int_{z}^{-\infty} x f(x, xv)\, dv\, dx + \int_{0}^{\infty} \int_{-\infty}^{z} x f(x, xv)\, dv\, dx \\
&= \int_{-\infty}^{0} \int_{-\infty}^{z} (-x) f(x, xv)\, dv\, dx + \int_{0}^{\infty} \int_{-\infty}^{z} x f(x, xv)\, dv\, dx \\
&= \int_{-\infty}^{z} \int_{-\infty}^{\infty} |x| f(x, xv)\, dx\, dv
\end{aligned}$$

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, xz)\, dx}$$

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz)\, dx}$$

If X, Y are independent.

Special case B: quotients

Example. How is the ratio of two independent Gaussians distributed?

Consider standard normal distribution, $Z=Y/X$, then

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{|x|}{2\pi} e^{-x^2/2} e^{-x^2 z^2/2}\, dx$$

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, xz)\, dx$$

Integrand is even,

$$f_Z(z) = \frac{1}{\pi} \int_0^{\infty} x e^{-x^2((z^2+1)/2)}\, dx$$

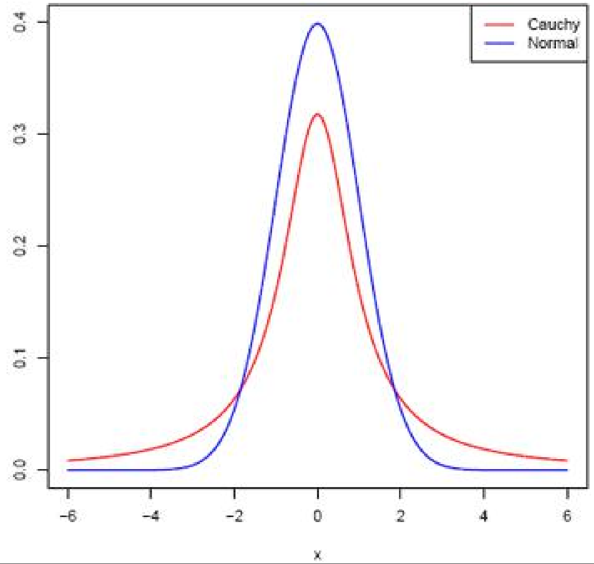
$$u = x^2$$

$$= \frac{1}{2\pi} \int_0^{\infty} e^{-u((z^2+1)/2)}\, du$$

$$\lambda = (z^2+1)/2$$

$$\int_0^{\infty} \lambda \exp(-\lambda x)\, dx = 1$$

$$f_Z(z) = \frac{1}{\pi(z^2+1)}, \qquad -\infty < z < \infty$$



Cauchy density decreases slower than Gaussians.

The general case

Example: *X*, *Y* are independent standard normal random variables, joint pdf is

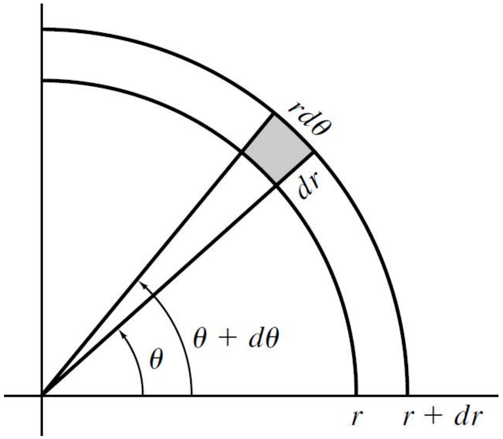
$$f_{XY}(x, y) = \frac{1}{2\pi} e^{-(x^2/2)-(y^2/2)}$$

Changing to polar coordinates,

$$R = \sqrt{X^2 + Y^2}$$

$$\Theta = \begin{cases} \tan^{-1}\left(\frac{Y}{X}\right), & \text{if } X > 0 \\ \tan^{-1}\left(\frac{Y}{X}\right) + \pi, & \text{if } X < 0 \\ \frac{\pi}{2}\operatorname{sgn}(Y), & \text{if } X = 0, Y \neq 0 \\ 0, & \text{if } X = 0, Y = 0 \end{cases}$$

$$X = R\cos\Theta$$
$$Y = R\sin\Theta$$



The general case

R, Θ have joint distribution

$$f_{R\Theta}(r,\theta)\,dr\,d\theta = P(r \le R \le r+dr, \theta \le \Theta \le \theta + d\theta)$$
$$= f_{XY}(r\cos\theta, r\sin\theta)r\,dr\,d\theta$$

$$f_{R\Theta}(r,\theta) = r f_{XY}(r\cos\theta, r\sin\theta)$$

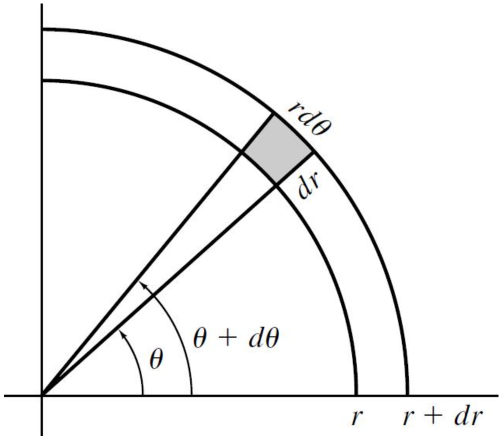
$$f_{R\Theta}(r,\theta) = \frac{r}{2\pi}e^{[-(r^2\cos^2\theta)/2-(r^2\sin^2\theta)/2]}$$
$$= \frac{1}{2\pi}re^{-r^2/2}$$

$$f_{XY}(x,y) = \frac{1}{2\pi}e^{-(x^2/2)-(y^2/2)}$$

Joint density implies that R and Θ are independent variables, Θ is uniform on $[0, 2\pi]$, R has the density

$$f_R(r) = re^{-r^2/2}, \qquad r \ge 0$$

Rayleigh density!



The general case: propositions

Two variables:

Under the assumptions just stated, the joint density of U and V is

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J^{-1}(h_1(u, v), h_2(u, v))|$$

for (u, v) such that $u = g_1(x, y)$ and $v = g_2(x, y)$ for some (x, y) and 0 elsewhere. ■

The general case:

If $X_1, \ldots, X_n$ have the joint density function $f_{X_1 \cdots X_n}$ and

$$Y_i = g_i(X_1, \ldots, X_n), \qquad i = 1, \ldots, n$$

$$X_i = h_i(Y_1, \ldots, Y_n), \qquad i = 1, \ldots, n$$

and if $J(x_1, \ldots, x_n)$ is the determinant of the matrix with the ij entry $\partial g_i / \partial x_j$, then the joint density of $Y_1, \ldots, Y_n$ is

$$f_{Y_1 \cdots Y_n}(y_1, \ldots, y_n) = f_{X_1 \cdots X_n}(x_1, \ldots, x_n)|J^{-1}(x_1, \ldots, x_n)|$$

wherein each x_i is expressed in terms of the y's; $x_i = h_i(y_1, \ldots, y_n)$.

Redo the previous example

X, Y are independent standard normal random variables, joint pdf is

$$f_{XY}(x, y) = \frac{1}{2\pi} e^{-(x^2/2)-(y^2/2)}$$

The roles of *u, v* are played by *r, θ*:

$$r = \sqrt{x^2 + y^2} \qquad \theta = \tan^{-1}\left(\frac{y}{x}\right) \quad \Longrightarrow \quad x = r\cos\theta \qquad y = r\sin\theta$$

Partial derivatives:

$$\frac{\partial r}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}} \qquad \frac{\partial r}{\partial y} = \frac{y}{\sqrt{x^2 + y^2}}$$

$$\frac{\partial \theta}{\partial x} = \frac{-y}{x^2 + y^2} \qquad \frac{\partial \theta}{\partial y} = \frac{x}{x^2 + y^2}$$

$$\Longrightarrow \quad J(x, y) = \frac{1}{\sqrt{x^2 + y^2}} = \frac{1}{r}$$

Proposition says:

$$f_{R\Theta}(r, \theta) = r f_{XY}(r\cos\theta, r\sin\theta)$$

for $r \geq 0$, $0 \leq \theta \leq 2\pi$, and 0 elsewhere

Lecture 4

Expected values

The expected value is a weighted average: the possible values are weighted by their probabilities. Discrete case:

DEFINITION

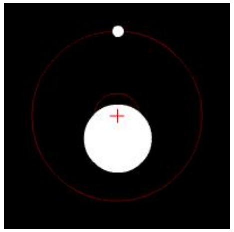
If X is a discrete random variable with frequency function $p(x)$, the expected value of X, denoted by $E(X)$, is

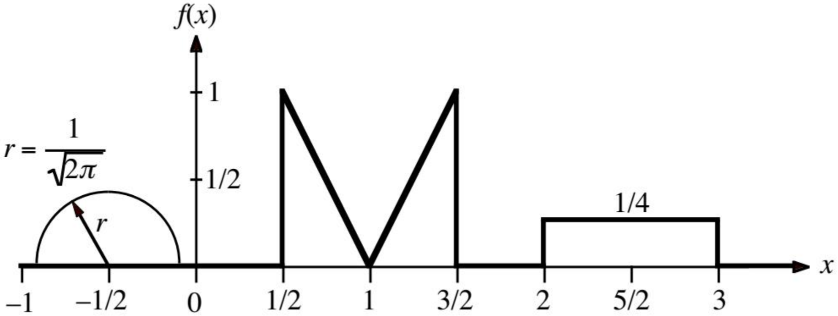
$$E(X) = \sum_i x_i p(x_i)$$

The expectation may be undefined!

$E(X)$ is also referred to as the **mean** of X and is often denoted by μ or μ_X.

Mean = Center of mass of the frequency function





Suppose that items produced in a plant are independently defective with probability p. Items are inspected one by one until a defective item is found. On the average, how many items must be inspected?

Make a guess!

Suppose that items produced in a plant are independently defective with probability p. Items are inspected one by one until a defective item is found. On the average, how many items must be inspected?

X, # of items expected, is a geometric random variable. $P(X=k)=q^{k-1}p,\ q=1-p.$

$$E(X)=\sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}$$

$$kq^{k-1}=\frac{d}{dq}q^{k}$$

$$= p\frac{d}{dq}\sum_{k=1}^{\infty} q^{k} = p\frac{d}{dq}\frac{q}{1-q}$$

$$= \frac{p}{(1-q)^2} = \frac{1}{p}$$

Suppose that items produced in a plant are independently defective with probability p. Items are inspected one by one until a defective item is found. On the average, how many items must be inspected?

X, # of items expected, is a geometric random variable. $P(X=k)=q^{k-1}p$, $q=1-p$.

$$E(X) = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}$$

$$kq^{k-1} = \frac{d}{dq}q^k$$

$$= p\frac{d}{dq}\sum_{k=1}^{\infty} q^k = p\frac{d}{dq}\frac{q}{1-q}$$

$$= \frac{p}{(1-q)^2} = \frac{1}{p}$$

$$E(X) = \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!}e^{-\lambda}$$

Poisson distribution has $E(X) = \lambda$

$$= \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}$$

An expected values=possible values weighted by probabilities. <u>Continuous case</u>:

DEFINITION

If X is a continuous random variable with density $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

provided that $\int |x|\, f(x) dx < \infty$. If the integral diverges, the expectation is undefined. ■

Example: Cauchy density has no expectation.

$$f(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right), \qquad -\infty < x < \infty \qquad\qquad \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} = \infty$$

An expected values=possible values weighted by probabilities. Continuous case:

DEFINITION

If X is a continuous random variable with density $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

provided that $\int |x|\, f(x) dx < \infty$. If the integral diverges, the expectation is undefined. ■

Example: Cauchy density has no expectation.

$$f(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right), \qquad -\infty < x < \infty \qquad \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} = \infty$$

Example: St. Petersburg Paradox

只赚不赔的赌博策略：**加倍下注法**。假设输赢概率各一半，第0次，下注1元；输了下2元，再输下4元，再输下8元……直到第k次赢了为止。最终输掉 $1+2+4+\ldots+2^{k-1}=2^k-1$元, 赢回$2^k$元，净得至少1元（实际赌博赢时不只回本）。

X=最后获胜一注数额，

$$P(X=2^k) = \frac{1}{2^{k+1}}$$

$$E(X) = \sum_{n=0}^{\infty} nP(X=n) = \sum_{k=0}^{\infty} 2^k \frac{1}{2^{k+1}} = \infty$$

Daniel Bernoulli (1700-1782)

Markov's Inequality

If X is a random variable with $P(X \geq 0) = 1$ and for which $E(X)$ exists, then $P(X \geq t) \leq E(X)/t$.

Discrete case:

$$E(X) = \sum_x xp(x) = \sum_{x<t} xp(x) + \sum_{x \geqslant t} xp(x)$$

All terms are nonnegative

$$\geqslant \sum_{x \geqslant t} xp(x) \geqslant \sum_{x \geqslant t} tp(x) = tP(X \geqslant t)$$

Probability that X is much bigger than $E(X)$ is small.

Let $t = k\,E(X)$, then $P\,(X > k\,E(X)) \leq 1/k$

对任意非负概率分布，大于几倍均值的概率至多是几分之一。

Expectations of functions of random variables

- If $Y=g(X)$, then $$E(Y) = \sum_x g(x)p(x)$$ $$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

- If $X_1,\ldots, X_n$ are jointly distributed random variables, $Y=g(X_1,\ldots, X_n)$, then

$$E(Y) = \sum_{x_1,\ldots,x_n} g(x_1,\ldots,x_n)p(x_1,\ldots,x_n)$$

$$E(Y) = \int\int\cdots\int g(x_1,\ldots,x_n)f(x_1,\ldots,x_n)dx_1\cdots dx_n$$

Question: $E[g(X)] = g[E(X)]$?

Example 1: The mean velocity is v_0, mean kinetic energy is $mv_0^2/2$?

Example 2: Constant voltage $V=IR$, measure I many times, average value is ~$E(I)$, $R=V/E(I)$?

Expectations of functions of random variables

- If $Y=g(X)$, then $$E(Y) = \sum_x g(x)p(x)$$ $$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

- If $X_1,\ldots, X_n$ are jointly distributed random variables, $Y=g(X_1,\ldots, X_n)$, then

$$E(Y) = \sum_{x_1,\ldots,x_n} g(x_1,\ldots,x_n)p(x_1,\ldots,x_n)$$

$$E(Y) = \int\int\cdots\int g(x_1,\ldots,x_n)f(x_1,\ldots,x_n)dx_1\cdots dx_n$$

Question: $E[g(X)] = g[E(X)]$? -- **No!**

Example 1: The mean velocity is v_0, mean kinetic energy is $mv_0^2/2$? -- **Roughly**

Example 2: Constant voltage $V=IR$, measure I many times, average value is $\sim E(I)$, $R=V/E(I)$? -- **Roughly**

Example: According to the kinetic theory of gases, the magnitude of the velocity of a gas molecule is random, following the Maxwell's distribution

$$f_X(x) = \frac{\sqrt{2/\pi}}{\sigma^3} x^2 e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$$

What is the mean kinetic energy $Y=mX^2/2$?

1. Find the pdf of Y, f_Y; then calculate $E(Y)$.

2. Use our theorem!

$$E(Y) = \int_0^\infty \frac{1}{2} m x^2 f_X(x)\, dx$$

$$= \frac{m}{2} \frac{\sqrt{2/\pi}}{\sigma^3} \int_0^\infty x^4 e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}\, dx$$

Change variable $u = x^2/2\sigma^2$

$$\frac{2m\sigma^2}{\sqrt{\pi}} \int_0^\infty u^{3/2} e^{-u}\, du = \frac{2m\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right)$$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

$$E(Y) = \tfrac{3}{2} m\sigma^2$$

Example: A stick of unit length is broken randomly in two places. What is the average length of the middle piece?

Make a guess!!

Example: A stick of unit length is broken randomly in two places. What is the average length of the middle piece?

Make a guess!!

The two break points are independent uniform random variables U_1 and U_2, find $E\,|U_1 - U_2|$, Now that $f(u_1, u_2) = 1$, $0 \le u_1 \le 1$, $0 \le u_2 \le 1$),

$$E|U_1 - U_2| = \int_0^1 \int_0^1 |u_1 - u_2|\, du_1\, du_2$$

$$= \int_0^1 \int_0^{u_1} (u_1 - u_2)\, du_2\, du_1 + \int_0^1 \int_{u_1}^1 (u_2 - u_1)\, du_2\, du_1$$

$u_2 \le u_1$ $\qquad$ $u_2 > u_1$

$= 1/3.$

Independence:

COROLLARY **A**

If X and Y are independent random variables and g and h are fixed functions, then $E[g(X)h(Y)] = \{E[g(X)]\}\{E[h(Y)]\}$, provided that the expectations on the right-hand side exist. ■

In particular, ***E*(*X*,*Y*) = *E*(*X*) *E*(*Y*) for independent *X*, *Y***.

Linearity:

If $X_1, \ldots, X_n$ are jointly distributed random variables with expectations $E(X_i)$ and Y is a linear function of the X_i, $Y = a + \sum_{i=1}^{n} b_i X_i$, then

$$E(Y) = a + \sum_{i=1}^{n} b_i E(X_i)$$

Example:
Expectation of binomial distribution is hard to evaluate directly.

$$E(Y) = \sum_{k=0}^{n} \binom{n}{k} k p^k (1-p)^{n-k}$$

Alternatively, consider Y as the sum of Bernoulli random variables, X_i=1 or 0 (success or failure on the ith trial).

$$Y = \sum_{i=1}^{n} X_i$$

$$E(X_i) = 0 \times (1-p) + 1 \times p = p \quad \Longrightarrow \quad E(Y) = np.$$

Variance and standard deviation

Variance and standard deviation

$$\mathrm{Var}(X) = E\{[X - E(X)]^2\}$$

Discrete

$$\mathrm{Var}(X) = \sum_i (x_i - \mu)^2 p(x_i)$$

Continuous

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

$$\sigma_Y = |b|\sigma_X$$

If $\mathrm{Var}(X)$ exists and $Y = a + bX$, then $\mathrm{Var}(Y) = b^2\mathrm{Var}(X)$.

Proof

Since $E(Y) = a + bE(X)$,

$$\begin{aligned} E[(Y - E(Y))^2] &= E\{[a + bX - a - bE(X)]^2\} \\ &= E\{b^2[X - E(X)]^2\} \\ &= b^2 E\{[X - E(X)]^2\} \\ &= b^2\mathrm{Var}(X) \end{aligned}$$

Variance and standard deviation

Bernoulli distribution

X takes on values 0 and 1 with probability $1-p$ and p, respectively.

Jacob Bernoulli
(1654–1705)

$$E(X_i) = 0 \times (1-p) + 1 \times p = p$$

$$\begin{aligned}\mathrm{Var}(X) &= (0-p)^2 \times (1-p) + (1-p)^2 \times p \\ &= p^2 - p^3 + p - 2p^2 + p^3 \\ &= p(1-p)\end{aligned}$$

Variance and standard deviation

Bernoulli distribution

X takes on values 0 and 1 with probability $1-p$ and p, respectively.

Jacob Bernoulli
(1654–1705)

$$E(X_i) = 0 \times (1-p) + 1 \times p = p$$

$$\begin{aligned} \mathrm{Var}(X) &= (0-p)^2 \times (1-p) + (1-p)^2 \times p \\ &= p^2 - p^3 + p - 2p^2 + p^3 \\ &= p(1-p) \end{aligned}$$

- Jacob Bernoulli (1654 - 1705; also known as James or Jacques) mathematician after whom Bernoulli numbers are named, and author of the early probability text
- Nicolaus Bernoulli (1662 - 1716) painter and alderman of Basel
- Johann Bernoulli (1667 - 1748; also known as Jean) mathematician and early adopter of infinitesimal calculus
- Nicolaus I Bernoulli (1687 - 1759) mathematician;
- Nicolaus II Bernoulli (1695 - 1726) mathematician; worked on curves, differential equations, and probability, and originator of the St. Petersburg paradox
- Daniel Bernoulli (1700 - 1782) developer of Bernoulli's principle and originator of the concept of expected utility for resolving the St. Petersburg paradox
- Johann II Bernoulli (1710 - 1790; also known as Jean) mathematician and physicist
- Johann III Bernoulli (1744 - 1807; also known as Jean) astronomer, geographer, and mathematician
- Jacob II Bernoulli (1759 - 1789; also known as Jacques) physicist and mathematician
- Hans Bernoulli, (1876 - 1959) architect, designer of the Bernoullihäuser in Zurich and Grenchen SO

$$\boxed{\mathrm{Var}(X) = E(X^2) - [E(X)]^2}$$

Proof: $\mathbf{Var}(X) = E[(X-\mu)^2] = (X^2 - 2\mu X + \mu^2)$

$$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

Example: Uniform distribution on [0, 1], $E(X)$=1/2,

$$E(X^2) = \int_0^1 x^2\,dx = \frac{1}{3}$$

$$\mathrm{Var}(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

Chebyshev's Inequality

Let X be a random variable with mean μ and variance σ^2. Then, for any $t > 0$,

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Proof: Let $Y=(X-\mu)^2$, then $E(Y)=\sigma^2$. Change t to t^2, apply Markov's inequality to Y.

$$P(X \geqslant t) \leqslant E(X)/t$$

Probability that X deviates much from $E(X)$ is small.

$$\boxed{\textbf{Let } t = k\sigma\textbf{, then } P(|X-\mu| \geq k\sigma) \leq 1/k^2}$$

对任意概率分布，偏离均值几σ的概率至多是几的平方分之一。

Example: Gaussian distribution, 2σ corresponds to how many %?

Here $P(|X-\mu| \geq 2\sigma) \leq 1/2^2 = 25\%$

Measurement error

Random error: a sequence of repeated independent measurements made with no deliberate change in the apparatus or experimental procedure still yield uncontrollable fluctuations, which are often modeled as random.

Systematic error: same effect on every measurement, e.g. equipment may be out of calibration, there may be errors associated with the measurement method.

$$\underbrace{X}_{\text{measurement}} = \underbrace{x_0}_{\text{true value}} + \overbrace{\beta}^{\text{systematic error / }\textbf{\textit{bias}}} + \underbrace{\varepsilon}_{\text{random error}}$$

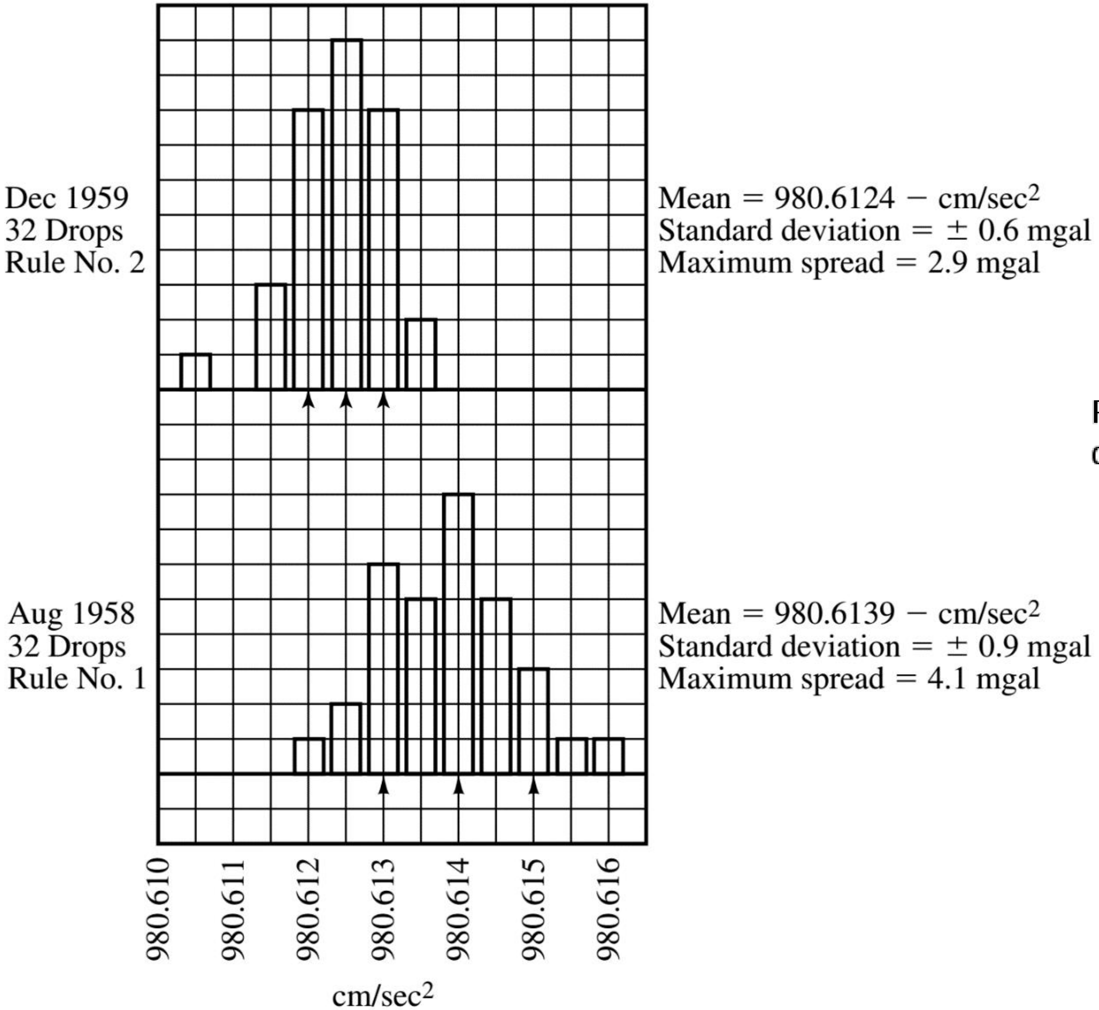
$$E(\varepsilon) = 0 \text{ and } \mathrm{Var}(\varepsilon) = \sigma^2$$

$$E(X) = x_0 + \beta$$

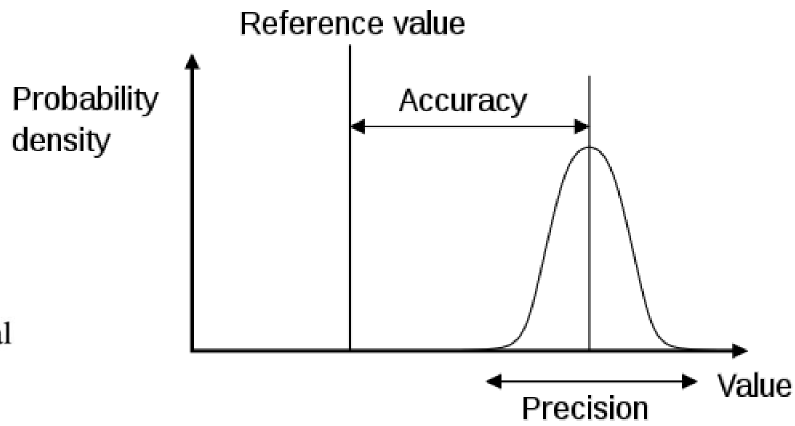
$$\mathrm{Var}(X) = \sigma^2$$

Measurement error

Acceleration due to gravity: Youden (1972), a NIST statistician. Measured at Ottawa, 32 times with each of two methods. Method #2 has smaller scatter.

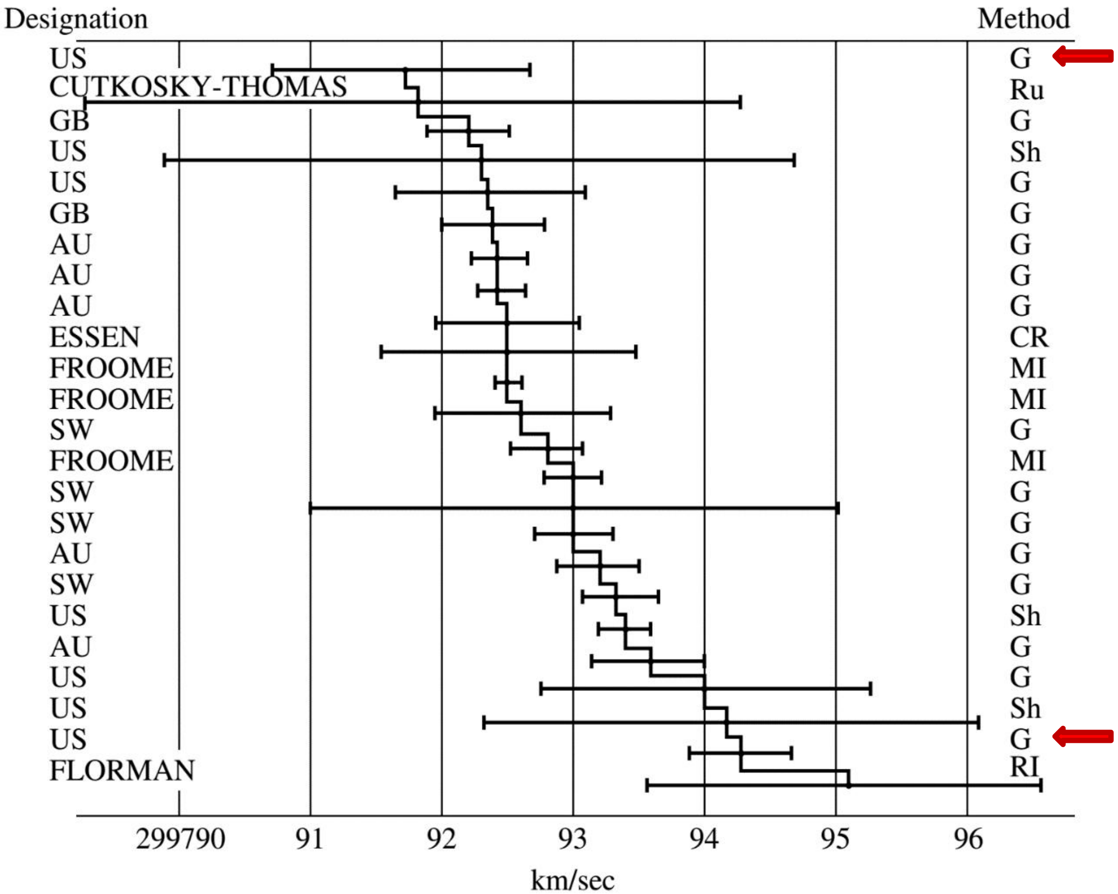


Precision vs. Accuracy



Measurement error

Speed of light: McNish (1962), Youden (1972).
24 independent determinations of *c*. Methods, e.g. G=geodimeter (光电测距仪）



1. Errorbars too small.
2. Spread of values cannot be accounted for by different experimental techniques alone.

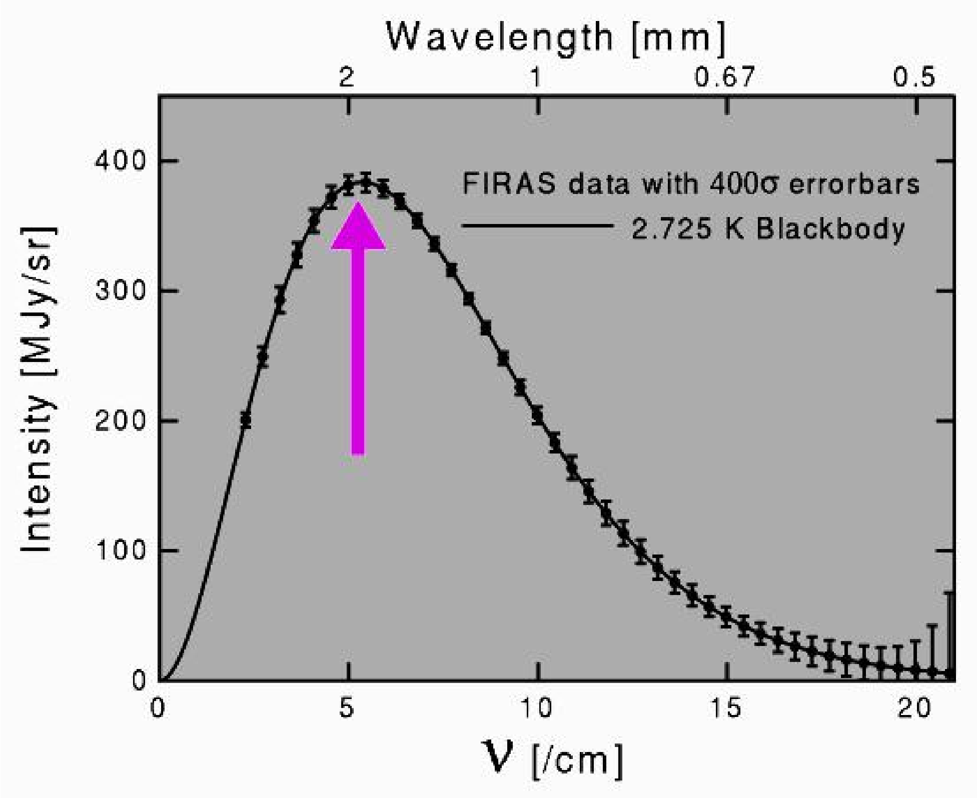
"Surely the evidence suggests that individual investigators are unable to set realistic limits of error to their reported values."

Measurement error

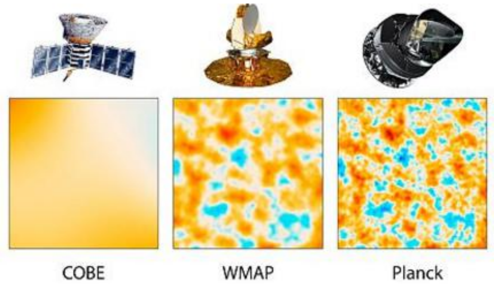
***What does it mean by, e.g.* 10.2±1.6?**

It is often not clear what precisely is meant by such notation. 10.2 is the experimentally determined value and 1.6 is *some measure* of the error.

It is often claimed or hoped that β is negligible relative to σ, and in that case 1.6 represents σ or some multiple of σ.



FIRAS instrument on COBE satellite, Nobel prize 2006



Measurement error

***What does it mean by, e.g.* 10.2±1.6?**

It is often not clear what precisely is meant by such notation. 10.2 is the experimentally determined value and 1.6 is *some measure* of the error.

It is often claimed or hoped that β is negligible relative to σ, and in that case 1.6 represents σ or some multiple of σ.

An overall measure of the size of the measurement error that is often used is the **mean squared error** $\mathrm{MSE} = E[(X - x_0)^2]$

$$\boxed{\mathrm{MSE} = \beta^2 + \sigma^2}$$

$$\begin{aligned} E[(X - x_0)^2] &= \mathrm{Var}(X - x_0) + [E(X - x_0)]^2 \\ &= \mathrm{Var}(X) + \beta^2 \\ &= \sigma^2 + \beta^2 \end{aligned}$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

Covariance and correlation

Covariance (协方差)

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Positively associated: $X >$ mean, $Y >$ mean, Cov > 0;
Negatively associated: $X >$ mean, $Y <$ mean, Cov < 0.

$$\begin{aligned}\text{Cov}(X, Y) &= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y) \\ &= E(XY) - E(X)\mu_Y - E(Y)\mu_X + \mu_X\mu_Y\end{aligned}$$

$$\boxed{\textbf{Cov } (X, Y) = E\ (XY) - E(X)\ E(Y)}$$

X, Y independent, $E(XY)=E(X)E(Y)$, Cov $(X, Y)=0$

In parallel, $$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Bilinear property of covariance

$U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{m} d_j Y_j$. Then

$$\mathrm{Cov}(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j \mathrm{Cov}(X_i, Y_j)$$

Therefore:

$$\mathrm{Var}(a + \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \mathrm{Cov}(X_i, X_j).$$

$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$, if the X_i are independent.

$$\boxed{\mathbf{Var}\ (X+Y) = \mathbf{Var}\ (X) + \mathbf{Var}\ (Y) + 2\mathbf{Cov}\ (X, Y)}$$

独立与否，均有 $E(\sum X_i) = \sum E(X_i)$

只有独立，才有 $\mathrm{Var}(\sum X_i) = \sum \mathrm{Var}(X_i)$

Example: Random Walk

A drunken walker starts out at a point x_0 on the real line. He takes a step on length X_1, which is a random variable with expected value μ and variance σ, and his position at that time is $S(1) = x_0 + X_1$. He then takes another step of length X_2, which is independent of X_1 with the same mean and standard deviation. His position after n such steps is $S(n) = x_0 + \sum_{i=1}^{n} X_i$. Then

$$E(S(n)) = x_0 + E\left(\sum_{i=1}^{n} X_i\right) = x_0 + n\mu$$

$$\mathrm{Var}(S(n)) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2$$

Brownian motion is a continuous time version of a random walk with the steps being normally distributed random variables.

- Robert Brown: 1827, apparently spontaneous motion of pollen grains suspended in water
- Albert Einstein: 1905, due to collisions with randomly moving water molecules
- Louis Bachelier: 1900, PhD thesis “The theory of speculation” related random walks to the evolution of stock prices

Correlation coefficient (相关系数)

X and Y are jointly distributed random variables,

$$\rho = \frac{\mathrm{Cov}(X,\ Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \qquad \text{or} \qquad \rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

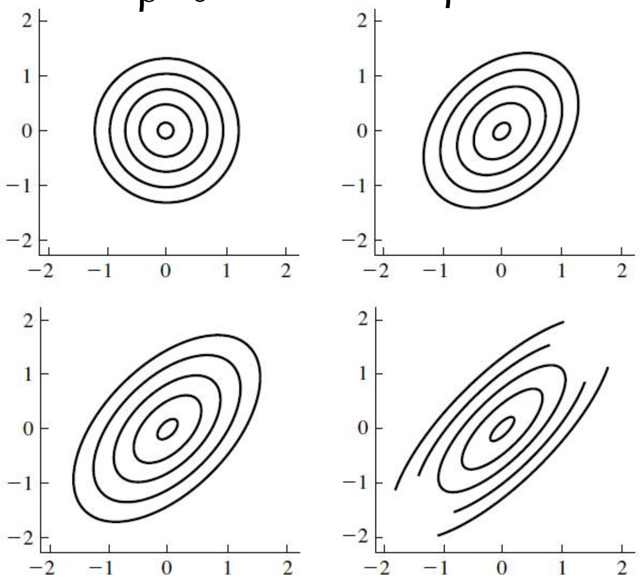
Example.

X, Y follow a bivariate normal distribution, their covariance is $\rho\sigma_X\sigma_Y$

$\mu_X = \mu_Y = 0$

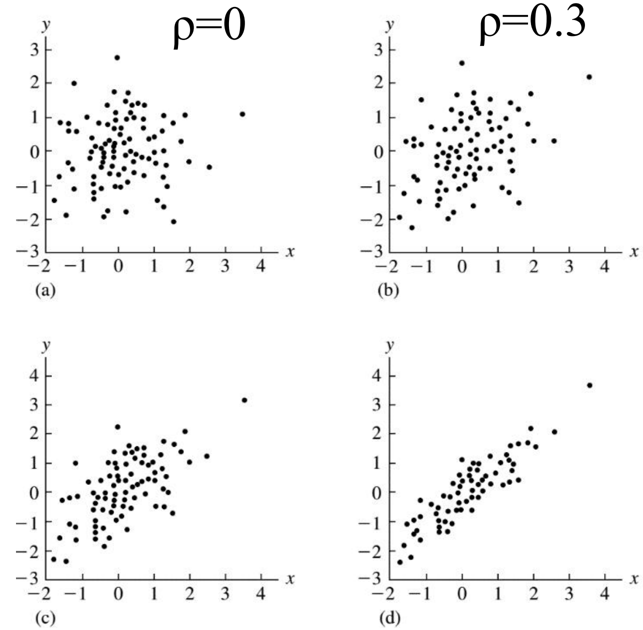
$\sigma_X = \sigma_Y = 1$

ρ=0 ρ=0.3



ρ=0.6 ρ=0.9

ρ=0 ρ=0.3



ρ=0.6 ρ=0.9

Problem set #2

Suppose that in a numerical simulation, you need to generate some fake noise that follows a **standard normal distribution**. Since the cdf has no closed form, let us do it in two ways:

1. *Rejection method*. (cf. Lec 3, pp. 28). Although a Gaussian is defined on $(-\infty, \infty)$, you can approximate it by a large enough interval (e.g. $[-3, 3]$, $[-5, 5]$), and choose $m(x)$ to be uniformly distributed. Plot a histogram using your output data.

2. *Polar method.* In Lec 4 we find that if X, Y are Gaussian, Θ is uniform on $[0, 2\pi]$, R has a Rayleigh density. Let $T=R^2$, by calculating the cdf (Lec 2, pp. 32), we obtain

$$f_T(t) = \frac{1}{2}e^{-t/2}, \qquad t \geq 0 \qquad\qquad f_{T\Theta}(t, \theta) = \frac{1}{2\pi}\left(\frac{1}{2}\right)e^{-t/2}$$

where the joint density is due to the fact that T, Θ are also independent. R^2 is exponential!

Therefore, we can the do following: First, you generate independent random variables U_1 and U_2, both uniformly distributed on $[0, 1]$. Then $-2 \log U_1$ is exponentially distributed with parameter ½, and $2\pi\, U_2$ is uniform on $[0, 2\pi]$. Therefore, the following X, Y are independent standard normal random variables. Again, plot a histogram demonstrating that your output indeed follows $N(0, 1)$.

$$X = \sqrt{-2\log U_1}\cos(2\pi U_2) \qquad\qquad Y = \sqrt{-2\log U_1}\sin(2\pi U_2)$$