Chapter 23

~ Probability and Statistics

Probability

$\boxed{\text{mutually exclusive events}}$

- sample space S

Probability

$$P(x_i) := \lim_{n \to +\infty} \frac{\text{number of } x_i \text{ occurs}}{\text{total trails}}$$

Theoretical probability

$$P(x_i) := \frac{\text{number of outcomes } x_i}{\text{total number of all events}}$$

example.

1. Tossing Coins . 2 .
2. sand

Axioms.

1. $0 \leq P \leq 1$
2. $\sum_i P_i = 1$
3. Probability of mutually exclusive events add.

Conditional Probability

$$P(A, B) = P(A) \cdot P(B \mid A).$$

$$P(B \mid A) = \frac{P(A, B)}{P(A)}$$

If A, B independent,

$$P(A, B) = P(A)P(B).$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Bayes' theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

example.

accuracy 99%, morbidity 0.5%.

A : sick

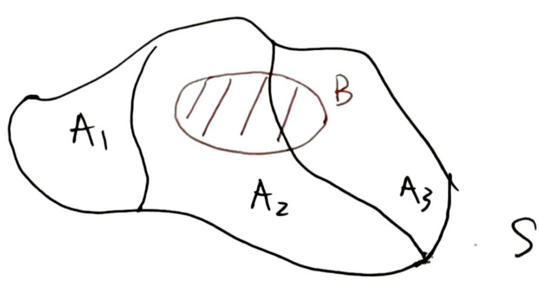
B : diagnosed. $\sim 33.2\%$.

more general theorem.

A_i mutually exclusive and $\bigcup_i A_i = S$

$\forall B \subset S$, we have

$$P(B) = \sum_i P(A_i) P(B|A_i)$$



Permutations and Combinations

$$P_m^n \qquad A_m^n = \frac{m!}{(m-n)!}$$

$$C_m^n \qquad \binom{m}{n} = \frac{m!}{n!(m-n)!}$$

$$B(n_1, n_2 \cdots n_i) = \frac{n!}{n_1!\, n_2! \cdots}$$

multinomial coefficient

$$(x_1 + \cdots + x_m)^n \longrightarrow \boxed{}\, x_1^{n_1} x_2^{n_2} \cdots x_m^{n_m}$$

Random Variables

Mean, deviation, correlation.

probability density.

$$P(x \le X \le x + dx) = f(x)\, dx$$

$$f(x) \geq 0 \quad \& \quad \int f(x)\, dx = 1$$

example. $\psi = N e^{-\frac{r}{a}}$ is electron. a->Bohr radius

$$N = \frac{1}{\sqrt{\pi a^3}}$$

$$\int |\psi|^2 d^3 r = 4\pi N^2 \int_0^\infty e^{-\frac{2r}{a}} r^2 dr = \pi a^3 N^2 = 1$$

Mean and Variance

$$\bar{x} = \frac{1}{n} \sum_j x_j$$

$$\langle X \rangle = \sum_i x_i p_i$$

$$= \int x f(x)\, dx$$

Why not

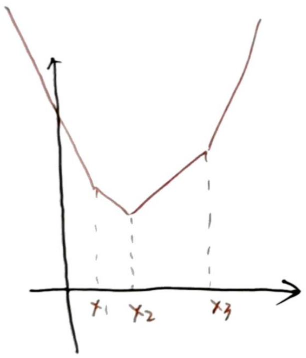
$$x_g = (x_1 x_2 \cdots x_n)^{1/n}$$

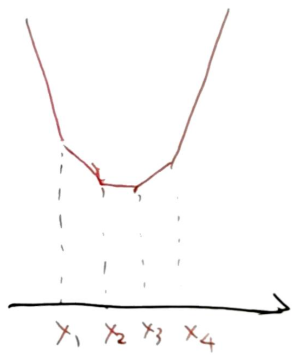
or

$$\frac{1}{x_h} = \frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n} \right)$$

or

$\tilde{x}$ minimize $\sum |x_i - x|$





minimize $\sum_i (x - x_i)^2$

$$\frac{\partial \sum_i (x - x_i)^2}{\partial x} = 2 \sum (x - x_i) = 0$$

$$\Rightarrow \quad n x = \sum x_i$$

$$x = \frac{1}{n} \sum_i x_i$$

method of least squares.

population deviation

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2}$$

squaring

$$n\sigma^2 = \sum_i x_i^2 - 2\langle x \rangle \sum_i x_i + n \langle x \rangle^2$$
$$= n\left(\langle x^2 \rangle - \langle x \rangle^2\right)$$

So,

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$$

variance.

$$\sigma^2 = \sum_j (x_j - \langle x \rangle)^2 p_j$$
$$= \int (x - \langle x \rangle)^2 f(x)\, dx$$

$$Y = aX + b.$$

Then $\langle Y \rangle = a \langle X \rangle + b$

$$\sigma^2(Y) = \int (ax + b - a\langle X \rangle - b)^2 f(x)\, dx$$
$$= \int a^2 (x - \langle x \rangle)^2 f(x)\, dx$$
$$= a^2 \sigma^2(X)$$

Moments of Probability Distributions

$$\langle (X - \langle X \rangle)^k \rangle = \sum_j (x_j - \langle X \rangle)^k p_j$$

$$= \int_{-\infty}^{+\infty} (x - \langle X \rangle)^k f(x)\, dx$$

moment-generating function

$$\langle e^{tX} \rangle = \int e^{tx} f(x)\, dx = 1 + t\langle X \rangle + \frac{t^2}{2!}\langle X^2 \rangle$$

Therefore.

$$\langle X \rangle = \frac{d\langle e^{tX} \rangle}{dt}\Big|_{t=0}$$

$$\langle X^2 \rangle = \frac{d^2 \langle e^{tX} \rangle}{dt^2}\Big|_{t=0}$$

$$\vdots$$

central moments.

example.

$\boxed{1}$ $\boxed{2}$ $\boxed{3}$ $\boxed{4}$

draw two, add.

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

$P(7) = \frac{1}{6}$.

$$M = \frac{1}{6}\left(e^{3t} + e^{4t} + 2e^{5t} + e^{6t} + e^{7t}\right)$$

$$M' = \frac{1}{6}\left(3e^{3t} + 4e^{4t} + 10e^{5t} + 6e^{6t} + 7e^{7t}\right)$$

$$M'' = \frac{1}{6}\left(9e^{3t} + 16e^{4t} + 50e^{5t} + 36e^{6t} + 49e^{7t}\right)$$

So,

$$\langle X \rangle = 5$$

$$\langle X^2 \rangle = \frac{80}{3}$$

$$\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2 = \frac{5}{3}$$

Covariance and Correlation.

$$f(x, y) = f(x)\, g(y) \Leftrightarrow \text{independent}.$$

$$\mathrm{cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle$$

if independent.

$$= \int (X - \langle X \rangle)(Y - \langle Y \rangle) f(x, y)\, dx\, dy$$

$$= \int (X - \langle X \rangle) f(x)\, dx \int (Y - \langle Y \rangle)\, g(y)\, dy$$

$$= 0$$

normalized covariance : correlation

$$-1 \leq \frac{\mathrm{cov}(X, Y)}{\sigma(X)\, \sigma(Y)} \leq 1$$

$$Q = \langle [a(X-\langle X\rangle) + c(Y-\langle Y\rangle)]^2 \rangle$$

$$= a^2 \langle (X-\langle X\rangle)^2 \rangle + 2ac \underbrace{\langle (X-\langle X\rangle)(Y-\langle Y\rangle) \rangle}_{\mathrm{cov}(X,Y)}$$

$$+ c^2 \langle (Y-\langle Y\rangle)^2 \rangle \geq 0$$

$$a^2 \sigma(X)^2 + 2ac\,\mathrm{cov}(X,Y) + c^2 \sigma(Y)^2 \geq 0$$

So, it's discriminant

$$4c^2 \mathrm{cov}(X,Y)^2 - 4c^2\sigma(X)^2 \sigma(Y)^2 \geq 0$$

$$\mathrm{cov}(X,Y)^2 \leq \sigma(X)^2 \sigma(Y)^2$$

$$Y = aX + b \Leftrightarrow \mathrm{cov}(X,Y)/\sigma(X)\sigma(Y) = \pm 1$$

Marginal Probability Distributions.

$$F(x) = \int f(x,y)\,dy$$

$$G(y) = \int f(x,y)\,dx.$$

↑

marginal.

Conditional Probability Distributions.

$$P(X = x \mid Y = y_0)$$

$$f(x, y_0)$$

Binomial Distribution.

repeated independent trials

example.

a toss. got 6, $p = \frac{1}{6}$

4 times.

$$P(S=s) = \binom{4}{s} a^s b^{4-s}$$

$$\sum_{s=0}^{4} \binom{4}{s} a^s b^{4-s} = 1.$$

$$\boxed{P(S=s) = \binom{n}{s} p^s q^{n-s}} \qquad q = 1-p$$

example. moment-generating function.

$$\langle e^{tS} \rangle = \langle e^{t(S_1 + \cdots + S_n)} \rangle$$

$$= \langle e^{tS_1} \rangle \langle e^{tS_2} \rangle \langle \cdots \rangle \cdots \langle e^{tS_n} \rangle$$

$$= (\langle e^{tS_1} \rangle)^n$$

$$\langle e^{tS_1} \rangle = pe^t + q$$

So,

$$\langle e^{tS} \rangle = (pe^t + q)^n$$

$$\frac{\partial \langle e^{tS} \rangle}{\partial t} = npe^t (pe^t + q)^{n-1}$$

$t = 0 \rightarrow \langle s \rangle = np$

$$\frac{\partial^2 \langle e^{ts} \rangle}{\partial t^2} = npe^t(pe^t + q)^{n-1} + n(n-1)p^2 e^{2t}(pe^t + q)^{n-2}$$

$$\langle s^2 \rangle = np + n(n-1)p^2$$

$$\sigma^2(s) = \langle s^2 \rangle - \langle s \rangle^2 = npq$$

Poisson Distribution.

example.

Decay , Poisson noise.

repeatedly at a constant rate of probability

$$P_n(t+dt) = P_n(t)\, P_0(dt) + P_{n-1}(t)\, P_1(dt)$$

$$P_1(dt) = \mu dt << 1 \qquad P_0(dt) = (1-\mu)dt$$

So,

$$\frac{dP_n(t)}{dt} = \frac{P_n(t+dt) - P_n(t)}{dt} = \mu P_{n-1}(t) - \mu P_n(t)$$

set $n = 0$

$$\frac{dP_0(t)}{dt} = \mu P_0(t) \Rightarrow P_0(t) = e^{-\mu t}$$

$$P_n(t) = \frac{(\mu t)^n}{n!} e^{-\mu t}$$

$$P_n(0) = \delta_{n0}.$$

$$\mu t \to \mu$$

$$p(n) = \frac{\mu^n}{n!} e^{-\mu}$$

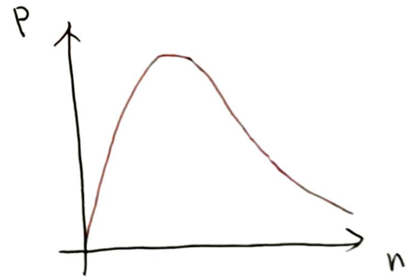
normalized.

$$\langle X \rangle = \sum_{n=1}^{\infty} n \frac{\mu^n}{n!} e^{-\mu} = e^{-\mu} \sum_{n=1}^{\infty} \frac{\mu^n}{(n-1)!} = \mu$$

$$\langle X^2 \rangle = \sum_{n=1}^{\infty} n^2 \frac{\mu^n}{n!} e^{-\mu} = e^{-\mu} \sum_{n=1}^{\infty} \left[\frac{\mu^n}{(n-2)!} + \frac{\mu^n}{(n-1)!} \right] = \mu^2 + \mu$$

$$\sigma^2 = \mu$$

$$\langle e^{tX} \rangle = \sum_{n=0}^{\infty} \frac{\mu^n}{n!} e^{-\mu} e^{tn} = e^{-\mu} \sum_{n=0}^{\infty} \frac{(\mu e^t)^n}{n!} = e^{\mu(e^t - 1)}$$



Relation to Binomial Distribution

$$P(S = s) = \frac{n!}{s!(n-s)!} p^s q^{n-s} \qquad \begin{array}{l} n \to +\infty \\ p \to 0 \end{array} \qquad np \to \mu$$

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

$$\frac{n!}{(n-s)!} \sim \left(\frac{n}{e}\right)^n \left(\frac{e}{n-s}\right)^{n-s} \sim \left(\frac{n}{e}\right)^s \left(\frac{n}{n-s}\right)^{n-s}$$

$$= \left(\frac{n}{e}\right)^s \left(1 + \frac{s}{n-s}\right)^{n-s}$$

So,

$$\frac{n!}{(n-s)!} \sim n^s$$

$$q^{n-s} = (1-p)^{n-s} \qquad , p = \frac{\mu}{n}$$

$$\sim \left(1-\frac{\mu}{n}\right)^n \left(1-\frac{\mu}{n}\right)^{-s}$$

$$\sim e^{-\mu}$$

So,
$$P(S=s) = \frac{n!}{s!(n-s)!} p^s q^{n-s}$$

$$\sim \frac{n^s}{s!} p^s e^{-\mu}$$

$$= \frac{\mu^s}{s!} e^{-\mu} .$$

Gauss' Normal Distribution.

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

limits of Poisson and Binomial Distributions.

$$\ln p(n) = \ln\left(\frac{\mu^n e^{-\mu}}{n!}\right) = n\ln\mu - \mu - \ln\sqrt{2n\pi} - n\ln n + n$$

$$\overset{n \to \mu+\nu}{=} (\mu+\nu)\ln\left(\frac{\mu}{\mu+\nu}\right) + \nu - \ln\sqrt{2(\mu+\nu)\pi}$$

$$= (\mu+\nu)\ln\left(1-\frac{\nu}{\mu+\nu}\right) + \nu - \ln\sqrt{2\pi(\mu+\nu)}$$

$$= -\sum_{t=1}^{\infty} \frac{\nu^t}{t(\mu+\nu)^{t-1}} + \nu - \ln\sqrt{2\pi(\mu+\nu)}$$

$$\ln p(n) \sim -\frac{v^2}{2\mu} - \ln\sqrt{2\pi\mu} \qquad (\mu \to +\infty)$$

$$\Leftrightarrow \quad p(n) \sim \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{v^2}{2\mu}}$$

$$\sigma = \sqrt{\mu}$$

Transformations of Random Variables.

$$X \to Y \qquad y = y(x)$$

$$g(y)\,dy = f(x)\,dx$$

$$X, Y \to U(X, Y), V(X, Y)$$

$$g(u, v) = f(x(u, v), y(u, v)) |J|$$

where $$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Addition.

$$x = x$$
$$z = x + y$$

$$J = 1$$

$$g(x, z) = f(x, z - x)$$

$$g(z) = \int_{-\infty}^{+\infty} f(x, z - x)\,dx \qquad \text{, independent.}$$

$$g(z) = \int_{-\infty}^{+\infty} f_1(x) f_2(z-x)\, dx$$
$$= \sqrt{2\pi}\, (f_1 * f_2)(z)$$

$$\langle e^{itx} \rangle = \int e^{itx} f(x)\, dx = \sqrt{2\pi}\, f^T(t)$$

$$g^T(t) = \sqrt{2\pi}\, f_1^T(t) f_2^T(t) .$$

$$g(z) = \int e^{-izt} f_1^T(t) f_2^T(t)\, dt$$

$$\langle e^{itz} \rangle = \langle e^{itX} \rangle \langle e^{itY} \rangle$$

Multiplication.

$$x = x \qquad J = \frac{1}{x}$$
$$z = xy$$

$$g(z) = \int_{-\infty}^{+\infty} f\left(x, \frac{z}{x}\right) \frac{dx}{|x|}$$
$$= \int f_1(x) f_2\left(\frac{z}{x}\right) \frac{dx}{x}$$

Gamma Distribution

$$Y = X^2 \qquad X \sim N(\mu, \sigma^2) .$$
$$\mu \downarrow 0$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} .$$

$$g(y) = \frac{e^{-\frac{y}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \cdot \frac{\sqrt{y}}{2} \qquad , y > 0 .$$

$$0 \qquad , y \le 0 .$$

$$z = \frac{y}{2\sigma^2}$$ normalized

$$\Rightarrow \int g(z)\,dz = \frac{1}{\sqrt{\pi}} \int_0^\infty z^{-\frac{1}{2}} e^{-z} dz = \frac{\Gamma(\frac{1}{2})}{\sqrt{\pi}} = 1$$

gamma distribution

$$g(p,\sigma;y) = \begin{cases} 0, & y \le 0 \\ \dfrac{y^{p-1} e^{-\frac{y}{2\sigma^2}}}{(2\sigma^2)^p \Gamma(p)}, & y > 0 \end{cases}$$

$$[g(p,\sigma)]^T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{(1-2i\sigma^2 t)^p}$$

$$\langle e^{itx} \rangle = \frac{1}{(1-2i\sigma^2 t)^p}$$

Statistics.

error propagation.

$$x_j = \bar{x} + e_j \qquad \sum_j e_j = 0$$

$$\bar{f} = \frac{1}{n}\sum f(x_j) = \frac{1}{n}\sum f(\bar{x} + e_j)$$

$$= f(\bar{x}) + \frac{1}{n} f'(\bar{x}) \sum e_j + \frac{1}{2n} f''(\bar{x}) \sum e_j^2$$

$$= f(\bar{x}) + \frac{1}{2}\sigma^2 f''(\bar{x}) + \cdots$$

$$\sigma^2(f) = \frac{1}{n}\sum_j (f_j - \bar{f})^2 \propto [f'(\bar{x})]^2 \cdot \frac{1}{n}\sum e_j^2$$

$$= [f'(\bar{x})]^2 \sigma^2$$

$$f(\bar{x} \pm \sigma) = f(\bar{x}) \pm f'(\bar{x})\sigma$$

two variant

$$\sigma^2(f) = \frac{1}{rs}\sum_j \sum_k (f_{jk} - \bar{f})^2$$

$$= \frac{1}{rs}\sum\sum (f_x u_j + f_y v_k)^2$$

$$= \frac{f_x^2}{r}\sum u_j^2 + \frac{f_y^2}{s}\sum v_k^2$$

So, $\sigma^2(f) = f_x^2 \sigma_x^2 + f_y^2 \sigma_y^2$.

Repeated Measurements.

$e_j = x_j - \bar{x}$

$\sum_j e_j = 0$

$$s^2 = \frac{1}{n}\sum_j v_j^2 = \frac{1}{n}\sum_{j=1}^{n}(e_j + \alpha)^2$$

$$= \frac{1}{n}\sum_j (e_j^2 + \alpha^2)$$

$$\alpha^2 \simeq \frac{s^2}{n}$$

$$\Rightarrow \quad s^2\left(1 - \frac{1}{n}\right) = \frac{1}{n}\sum e_j^2$$

$$s = \sqrt{\frac{\sum_j (x_j - \bar{x})^2}{n-1}}$$

sample standard deviation.

The χ^2 Distribution.

$$\chi^2 := \sum_{j=1}^{n} \left(\frac{u_j - u(t_j, a, \cdots)}{\sigma_j} \right)^2$$

$X_i \sim$ Gamma Distribution.

$$\chi^2 = \sum X_j^2$$

$$g(\chi^2 = y) = \frac{y^{\frac{n}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$$

$$P(\chi^2 > y_0) = \int_{y_0}^{\infty} g(y)\, dy$$

Student t Distribution. W. S. Gosset.

$$T = \frac{Y\sqrt{n}}{\sqrt{\frac{S}{n}}}$$

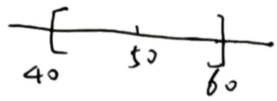
$$Y = \frac{1}{n} \sum_j x_j - \mu = \bar{x} - \mu$$

$$x_j \sim N$$

$$f_n(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\, \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

Confidence Interval

90%



$x = 0.50 \pm 0.10$ 90% confidence.

Gaussian

$$Y = \bar{X} - \mu = \frac{T\sqrt{s/n}}{\sqrt{n}}$$

$$P(-C_p < t < +C_p) = p$$

$$P(-\infty < t < +C_p) = \frac{1+p}{2} \equiv \hat{p}$$