For more than one variable

$$E_y[g(x,y)]$$

Normal distribution:

1st moment

$$E[x] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$$

Let $y = x-\mu \Rightarrow x = y+\mu \Rightarrow dy = dx$

$$E[x] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-y^2/2\sigma^2\right\} (y+\mu)\, dy$$

$$= \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2} y\, dy}_{\text{odd}} + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2} dy}_{1} = \mu$$

$$\mathrm{Var}[x] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} (x-\mu)^2\, dx$$

$y = x-\mu \qquad x = y+\mu \qquad dx = dy$

$$\mathrm{var}[x] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2} y^2\, dy = \int_{-\infty} \frac{(y)\left(y e^{-y^2/2\sigma^2}\right)}{\sqrt{2\pi}\,\sigma} dy$$

$$\boxed{\int uv = u\int v - \int\left(\int v\right) du}$$

$$= y \int_{-\infty}^{\infty} \frac{y e^{-y^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma} dy - \int\left[\int \frac{y e^{-y^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma} dy\right] dy$$

$$= y \cdot \int_{-\infty}^{\infty} \cancel{y}\, e^{-y^2/2\sigma^2} \left(-\frac{y}{\cancel{2}\sigma^2}\right)\left(-\cancel{2}\sigma^2\right) dy - \int\left[\frac{1}{\sqrt{2\pi}\,\sigma} \int e^{-y^2/2\sigma^2} \left(-\frac{y}{\cancel{2}\sigma^2}\right)\left(-\sigma^2\right) dy\right] dy$$

$$= \underbrace{y \cdot \frac{(-\sigma^2)}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2}\Big|_{-\infty}^{\infty}}_{\to 0} - \int_{-\infty}^{\infty} \frac{(-\sigma^2)}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2}\, dy$$

$$= \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2}\, dy = \sigma^2$$

Similarly 4th central moment $= 3\sigma^4$ (homework).

$\downarrow$

kurtosis $= 3$

If kurtosis (of other dist) > 3 [dist is leptokurtic]

" " < 3 [dist is platykurtic].

Leptokurtic dist $\Rightarrow$ risk of extreme events is underestimated

platykurtic " $\Rightarrow$ Extreme events are rare.

Examples of Discrete probability Mass functions:

Uniform Dist: flipping coin, rolling a die

If K possible outcomes,

prob for each outcome $= \frac{1}{K}$

$\boxed{\text{Fig 2.5, page 21}}$

· ~~Binary~~ Boolean variables are used when two possible outcomes

· e.g., head/tail, yes/no, 0/1, true/false.

$$p(x=1) = \mu \qquad 0 \le \mu \le 1$$

so $\quad p(x=0) = 1-\mu$

<u>Bernoulli Dist:</u> $\quad x \sim \text{Bern}(\mu)$.

$$\text{Bern}(x|\mu) = \mu^{x}(1-\mu)^{1-x} \qquad x \in \{0,1\}.$$

$B(x=1) = \mu$

$B(x=0) = 1-\mu$.

- $\boxed{\text{Fig 2.6, page 22}}$

Expected value : $\quad E[x] = 1\cdot\mu + 0\cdot(1-\mu) = \mu$

$$E[x^2] = 1^2\cdot\mu + 0^2\cdot(1-\mu) = \mu$$

$$\text{var}[x] = E[x^2] - E[x]^2 = \mu - \mu^2 = \mu(1-\mu).$$

- Extension:- N experiments, how many successes $(x=1)$?

Let m successes.

$$N\cdot(N-1)\cdot \ldots\ldots (N-m+1).$$

N → possibilities for 1st success.

$(N-1)$ → possibilities for 2nd success

. . .

$(N-m+1)$ → possibilities for last success.

$N\cdot(N-1)\ldots\ldots(N-m+1)$ # of m permutations.

$$= \frac{N(N-1)\cdot\ldots\ldots 1}{m(m-1)\ldots 1} = \frac{(N-m)!}{m!} = \binom{N}{m}$$

$\binom{N}{m}$ is the coefficient of a^m in the expansion of $(1+a)^N$.

$\boxed{\text{Fig 2.7, page 23}}$

This leads to binomial distribution.

$$Bin(m|N,\mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}$$

Expected $\quad E[m] = \sum_{m=0}^{N} m \, Bin(m|N,\mu) = N\mu$

Proof: $\quad m = X_1 + X_2 + \cdots + X_n$

$$E[m] = \underbrace{E[X_1]}_{} + E[X_2] + \cdots + E[X_n] = \mu + \mu + \cdots + \mu = N\mu.$$

$$E[X_1] = \cancel{x}\, 1 \cdot \underbrace{p(X_1=1)}_{\mu} + 0 \cdot p(X_1=0) = \mu.$$

$$\begin{aligned} Var[m] &= Var[X_1^2 + \cdots + X_N^2] \\ &= Var[X_1^2] + \cdots + \cancel{V}ar[X_N^2] \\ &= \mu(1-\mu) + \cdots + \mu(1-\mu) = N\mu(1-\mu). \end{aligned}$$

How many attempts $m \geq 1$ are necessary to succeed?

Geometric Distribution $\quad m \sim Geo(\mu)$.

\# of failure b/f 1st success.

$$Geo(m|\mu) = \mu(1-\mu)^{m-1}$$

Expected value:

$$E[m] = \sum_{m=0}^{\infty} m\mu(1-\mu)^{m-1}$$

$$= \mu \sum_{m=0}^{\infty} m(1-\mu)^{m-1}$$

$$= \mu \cdot \frac{1}{\mu^2} = \frac{1}{\mu}.$$

$$\sum_{j=0}^{\infty} x^j = \frac{1}{1-x}$$

$$\sum_{j=0}^{\infty} j x^{j-1} = \frac{1}{(1-x)^2}$$

$$E[m^2] = \sum_{m=0}^{\infty} m^2 \mu (1-\mu)^{m-1} = \mu \sum_{m=0}^{\infty} m^2 (1-\mu)^{m-1}$$

$1-\mu = x$

$$= \mu \cdot \frac{1+x}{(1-x)^3} = \mu \cdot \frac{1+(1-\mu)}{\mu^3}$$

$$= \frac{2-\mu}{\mu^2}$$

$$Var[x] = E[m^2] - E[m]^2$$

$$= \frac{2-\mu}{\mu^2} - \frac{1}{\mu^2} = \frac{2-\mu-1}{\mu^2}$$

$$= \frac{1-\mu}{\mu^2}$$

$$\sum_{j=0}^{\infty} j x^j = \frac{x}{(1-x)^2}$$

Diff again

$$\sum_{j=0}^{\infty} j^2 x^{j-1} = \frac{d}{dx}\left[\frac{x}{(1-x)^2}\right]$$

$$= \frac{1+x}{(1-x)^3}$$

Instead of using prob. of success, the rate λ of events occuring over a certain interval can be used.

real valued $\lambda > 0$:   examples: phone calls / hour

cars passing a traffic checkpoint per day.

Poisson Dist: $\quad m \sim P_o(\lambda)$

$m \downarrow$ # events occuring in a given time interval.

$$P_o(m|\lambda) = \frac{\lambda^m e^{-\lambda}}{m!}$$

$$E[m] = \sum_{m=0}^{\infty} \frac{m \lambda^m e^{-\lambda}}{m!} = e^{-\lambda} \sum \frac{m \lambda^m}{m!}$$

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

$$= e^{-\lambda} . \lambda \sum_{m=0}^{\lambda} \frac{\lambda^m}{m!} = e^{-\lambda} . \lambda . e^{\lambda} = \lambda$$

$$E[m^2] = \sum_{m=0}^{\infty} \frac{m^2 \lambda^m e^{-\lambda}}{m!}$$

$$= e^{-\lambda} \sum_{m=0}^{\infty} \frac{m^2 \lambda^m}{m!}$$

$$= e^{-\lambda} . \lambda (e^{\lambda} + \lambda e^{\lambda})$$

$$= \lambda (1 + \lambda)$$

$$\text{var}[m] = E[m^2] - E[m]^2$$

$$= \lambda(1+\lambda) - \lambda^2$$

$$= \lambda + \lambda^2 - \lambda^2 = \lambda$$

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

Diff.

$$\sum_{k=0}^{\infty} \frac{k \lambda^{k-1}}{k!} = e^{\lambda}$$

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

Diff.

$$\sum_{k=0}^{\infty} \frac{k^2 \lambda^{k-1}}{k!} = \frac{d}{dx}(\lambda e^{\lambda})$$

$$\sum_{k=0}^{\infty} \frac{k^2 \lambda^k}{k!} = \lambda \frac{d}{dx}(\lambda e^{\lambda})$$

$$= \lambda [e^{\lambda} + \lambda e^{\lambda}]$$

Categorical distribution:

$\underline{x} = (x_1, x_2, \dots, x_K)^T$ K-dim vector

Outcome can be any of K possible, mutually distinct outcomes

e.g. roll of a die, $\overset{1}{(}\;,\overset{2}{}\;,\overset{3}{}\;,\overset{4}{}\;,\overset{5}{}\;,\overset{6}{}\;)$

For 1 : $(1,0,0,0,0,0) \sim (x_1, x_2, x_3, x_4, x_5, x_6)$.
2 : $(0,1,0,0,0,0)$
3 : $(0,0,1,0,0,0)$
$\vdots$

Used mostly in categorical data.

Let mean of x_k is μ_k.

$\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_K)^T$

$$\mathrm{Cat}(\underline{x} \mid \underline{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

Only one is 1 others zero.

e.g 3 on die $\Rightarrow x_3 = 1$

$p(x_3 = 1 \mid \mu) = \mu_1^0 \cdot \mu_2^0 \cdot \mu_3^1 \cdot \mu_4^0 \cdot \mu_5^0 \cdot \mu_6^0$.
$= \mu_3$.

$p(x_k = 1 \mid \underline{\mu})$

generalization of Bernoulli dist

$$\sum_{\underline{x}} p(\underline{x} \mid \underline{\mu}) = \sum_{k=1}^{K} \mu_k = 1.$$

$$E[\underline{x} \mid \mu] = \sum_{\underline{x}} \underline{x} \cdot p(\underline{x} \mid \mu) = \underline{\mu}$$

$x_i \cdot p(x_i \mid \underline{\mu})$.

$\underset{0\;\;1}{\wedge}$ $p(x_i = 1 \mid \mu) = \mu_i$

$\mathrm{var}[x] = E[(x - E[x])^2] = E[x^2] - E[x]^2$ univariate

$\mathrm{var}[\underline{x}] = E[(\underline{x} - E(\underline{x}))(\underline{x} - E[\underline{x}])^T] = E[\underline{x}\underline{x}^T] - E[\underline{x}]E[\underline{x}]^T$

multivariate

So $\underline{x}\underline{x}^T$ ~~= I, identity matrix~~.

If $x_k = 1 \Rightarrow \underline{x}\underline{x}^T$ all zeros except (k,k) entry is 1.

e.g., $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Kth diagnal entry of var$[\underline{x}]$ is $\mu_k - \mu_k^2 = \mu_k(1-\mu_k)$.

Variance of Bernoulli dist.

The off diagnal entry (k, l) is known as covariance

$\text{cov}[x_k, x_l] = -\mu_k \mu_l$. (kth & lth component influence each other)

define $m_k = \sum_{n=1}^{N} x_{nk}$ (# of times $x_{nk} = 1$)

$\underline{m} = (m_1, \ldots, m_K)^T$ (it is actuall of length K-1)

$\sum_{k=1}^{K} m_k = N$ $\qquad m_K = N - m_1 - m_2 - \ldots - m_{K-1}$

The number of possibilities to generate a particular tuple is

$$\binom{N}{m_1 \ldots m_K} = \frac{N!}{m_1! \ldots m_K!}$$

[N balls, K bins, bin 1 has m_1 balls, bin 2 has m_2 balls .....]

[If K=2, then multinomial → binomial]

$\underline{m} \sim \text{Mult}(N, \underline{\mu})$ [Multinomial dist]

$$\text{Mult}(N, \underline{\mu}) = \binom{N}{m_1! \ldots m_K!} \prod_{k=1}^{K} \mu_k^{m_k}$$

$E[\underline{m}] = N\underline{\mu}$

$\underline{m} = \underline{X}_1 + \underline{X}_2 + \cdots + \underline{X}_n$

$E[\underline{m}] = E[\underline{X}_1] + E[\underline{X}_2] + \cdots + E[\underline{X}_n]$

$= \underline{\mu} + \underline{\mu} + \cdots + \underline{\mu} = N\underline{\mu}.$

Show: $var[\underline{m}] = N.\,diag(\underline{\mu}) - \underline{\mu}\underline{\mu}^T.$

Continuous PDF's

1. $x \sim Unif(a,b)$

$$Unif(x|a,b) = \begin{cases} \frac{1}{b-a} & a < x < b. \\ 0 & \text{otherwise}. \end{cases}$$

$$cdf: \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Expected: $E[x] = \int_a^b \frac{1}{b-a} x\, dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b = \frac{a+b}{2}$

$\cancel{Var[x]} \; E[x^2] = \int_a^b \frac{1}{b-a} x^2 dx = \frac{1}{b-a} \frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{3(b-a)}$

$= \frac{b^2+ab+a^2}{3}$

$Var[x]$

$= E[x^2] - E[x]^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4}$

$= \frac{4(b^2+ab+a^2) - 3(b^2+2ab+a^2)}{12} = \frac{b^2-2ab+a^2}{12} = \frac{(a-b)^2}{12}$

$x \sim N(\underline{\mu}, \underline{\Sigma})$

$$f(\underline{x}) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})\right]$$

* A random variable x, whose logarithm is normally distributed,

$y = \log x$ and $y \sim N(\mu, \sigma)$

$x \sim$ log normal distribution or Galton distribution.

$$\log N(x|\mu, \sigma^2) = \frac{1}{x\sqrt{2\pi}\,\sigma} e^{-(\log x - \mu)^2/2\sigma^2}$$

Expectation: Let $y = \log x$ $\qquad dy = \frac{1}{x} \cancel{\log x}\, dx$.

$$x\,dy = e^{y} dy$$

$$E[x] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2} e^{y} dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-((y-\mu)^2 + 2\sigma^2 y)/2\sigma^2} dy.$$

$$-(y-\mu)^2 + 2\sigma^2 y = -y^2 + 2\mu y - \mu^2 + 2\sigma^2 y$$

$$= -y^2 + 2(\mu+\sigma^2)y - \mu^2$$

$$= \underbrace{-y^2 + 2(\mu+\sigma^2)y - (\mu+\sigma^2)^2}_{-[y-(\mu+\sigma^2)]^2} + \underbrace{(\mu+\sigma^2)^2 - \mu^2}_{\cancel{\mu^2} + \sigma^4 + 2\mu\sigma^2 - \cancel{\mu^2}}$$

$$2\sigma^2\left(\frac{\sigma^2}{2} + \mu\right)$$

$$E[X] = e^{\mu+\frac{\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{[y-(\mu+\sigma^2)]^2}{2\sigma^2}\right\} dy$$

$$= e^{\mu+\sigma^2/2}$$

$E[X^2]$:

$$y = \log x \qquad dy = \frac{dx}{x}.$$

$$2y = \log x^2$$

$$e^{2y} = x^2$$

$$E[X^2] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2} \cdot e^{2y} dy.$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-[(y-\mu)^2+4\sigma^2 y]/2\sigma^2} dy.$$

$$-(y-\mu)^2+4\sigma^2 y = -y^2+2\mu y-\mu^2+4\sigma^2 y$$

$$= -y^2+2(\mu+2\sigma^2)y-\mu^2$$

$$= \underbrace{-y^2+2(\mu+2\sigma^2)y-(\mu+2\sigma^2)}_{-(y-(\mu+\sigma^2))^2}+\underbrace{(\mu+2\sigma^2)^2-\mu^2}_{\cancel{\mu^2}+4\sigma^4+4\mu\sigma^2-\cancel{\mu^2}}.$$

$$2\sigma^2(2\sigma^2+2\mu).$$

$$E[X^2] = e^{2\sigma^2+2\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{[y-(\mu+\sigma^2)]^2}{2\sigma^2}\right\} dy = e^{2\sigma^2+2\mu}$$

$$E[X^2]-E[X]^2 = e^{2\sigma^2+2\mu} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2}\left[e^{\sigma^2}-1\right].$$

If we know $E[x]$, ~~$E[x^2]$~~ $var[x]$ of log normal dist, mean and variance of associated normal dist. are

$$\mu = \log\left(\frac{E[x]^2}{\sqrt{Var[x] + E[x]^2}}\right)$$

$$\sigma^2 = \log\left(1 + \frac{Var[x]}{E[x]^2}\right)$$

$$var[x] = e^{2\mu+\sigma^2}\left[e^{\sigma^2} - 1\right].$$

$$= e^{2(\mu+\frac{\sigma^2}{2})}\cdot\left[e^{\sigma^2} - 1\right].$$

$E[x]$.

$$= E[x]^2\left(e^{\sigma^2} - 1\right).$$

$$\frac{Var[x]}{E[x]^2} = e^{\sigma^2} - 1 \Rightarrow e^{\sigma^2} = 1 + \frac{Var[x]}{E[x]^2}$$

$$\sigma^2 = \log\left(1 + \frac{Var[x]}{E[x]^2}\right)$$

$$e^{\sigma^2/2} = \sqrt{\frac{E[x]^2 + Var[x]}{E[x]^2}}$$

$$E[x] = e^{\mu+\sigma^2/2} = e^{\mu}\cdot e^{\sigma^2/2}$$

$$e^{\mu} = \frac{E[x]}{e^{\sigma^2/2}}$$

$$\mu = \log\frac{E[x]}{e^{\sigma^2/2}}$$

$$\mu = \log \frac{E[x]}{\sqrt{\frac{E[x]^2 + Var[x]}{E[x]^2}}}$$

$$= \log\left(\frac{E[x]^2}{\sqrt{E[x]^2 + Var[x]}}\right)$$