

L1. Probability review

Outline

- 1. Probability**
- 2. Discrete distributions**
- 3. Continuous distributions**

Outline

1. **Probability**
2. Discrete distributions
3. Continuous distributions

Probability as a measure of uncertainty

- Probability : A numerical measure of uncertainty

Y : Random variable

y : Data assigned to random variable Y

$$p_Y(Y = y) = p(y) \text{ :Short notation}$$

$p(y)$ if y is continuous, $p(y)$ is probability density function (PDF)
if y is discrete, $p(y)$ is probability mass function (PMF)

We are going to use $p(y)$ to represent both PDF and PMF

$Y \sim \text{dist}(\theta)$, θ is parameter for distribution

$$p_Y(Y = y|\theta) = p(y|\theta) = \text{dist}(y|\theta) \quad \text{or} \quad p_Y(Y = y; \theta) = p(y; \theta) = \text{dist}(y; \theta)$$

Example :

$$Y \sim N(\mu, \sigma^2) \rightarrow p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu-y)^2}{2\sigma^2}}$$

$$Y \sim \text{Bin}(n, \theta) \rightarrow p(y|\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

Useful results from Probability Theory for Bayesian Statistics

$$p(u, v)$$

$$p(u|v) = \frac{p(u, v)}{p(v)}$$

$$p(v) = \int_u p(u, v) du$$

$$p(u|v) = \frac{p(v|u)p(u)}{\int_u p(u, v) du}$$

$$p(u, v) = p(u|v) p(v)$$

$$p(u, v, w) = p(u|v, w)p(v|w)p(w)$$

$$p(u) = \int_v p(u|v)p(v) dv, \quad p(u) = \sum_v p(u|v)p(v)$$

Mean and Variance of conditional distributions

$$E(U) = \int up(u)du$$

$$E(g(U)) = \int g(u)p(u)du$$

$$\text{var}(U) = \int (u - E(U))^2 p(u)du$$

Conditional mean (law of total expectation)

$$E(U) = E_V(E(U|V))$$

$$\because E(U) = \int \int up(u, v)dudv = \int \int up(u|v)du p(v)dv = \int E(U|v)p(v)dv = E_V(E(U|V))$$

Mean and Variance of conditional distributions

Conditional variance (law of total variance)

- $\text{var}(U) = E[\text{var}(U|V)] + \text{var}(E[U|V])$

Proof:

$$\begin{aligned}\text{var}(U) &= E_U[U^2] - (E_U[U])^2 \\&= E_V[E_U[U^2|V]] - \{E_V[E_U[U|V]]\}^2 \quad \because \text{The law of total expectation} \\&= E_V[\text{var}(U|V) + (E_U[U|V])^2] - \{E_V[E_U[U|V]]\}^2 \quad \because \text{var}(X) = E(X^2) - (E(X))^2 \\&= E_V[\text{var}(U|V)] + E_V[(E_U[U|V])^2] - \{E_V[E_U[U|V]]\}^2 \quad \because \text{Linearity of expectation} \\&= E_V[\text{var}(U|V)] + E_V[(E_U[U|V])^2] - \{E_V[E_U[U|V]]\}^2 \\&= E_V[\text{var}(U|V)] + \text{var}(E_U[U|V])\end{aligned}$$

Example of Conditional Probability

$$p(u|v) = \frac{p(u, v)}{p(v)} = \frac{p(v|u)p(u)}{\int_u p(u, v) du}$$

Given probability

$$p(C) = \frac{1}{100} \rightarrow p(NC) = \frac{99}{100}$$

$$p(+|C) = \frac{90}{100} \rightarrow p(-|C) = \frac{10}{100}$$

$$p(+|NC) = \frac{8}{100} \rightarrow p(-|NC) = \frac{92}{100}$$

$p(C)$: Probability of having cancer

$p(+|C)$: Positive result given cancer

$p(+|NC)$: Positive result given no cancer

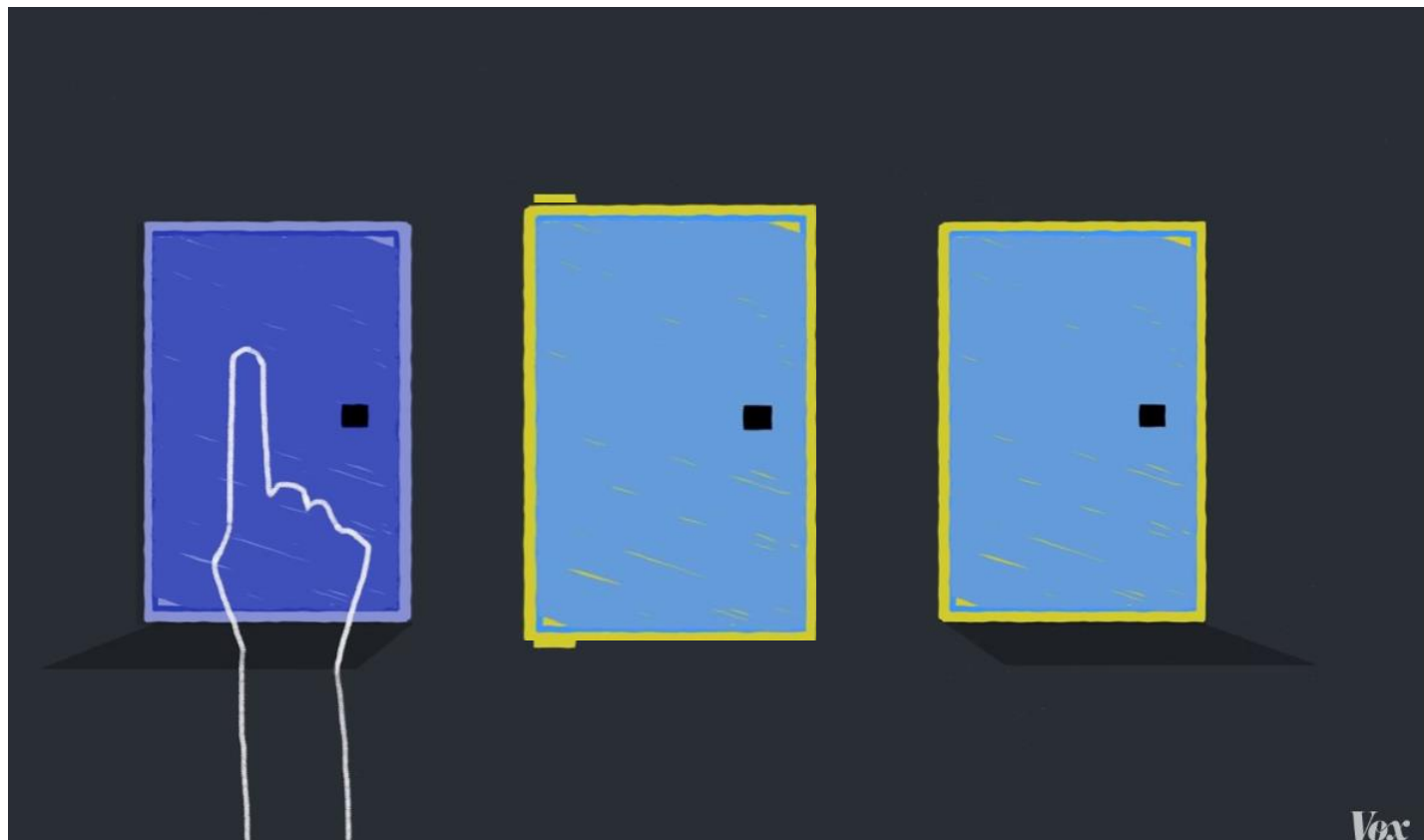
$$\begin{aligned} p(C|+) &= \frac{p(C, +)}{p(+)} = \frac{p(+|C)p(C)}{p(+)} \\ &= \frac{p(+|C)p(C)}{p(+|C)p(C) + p(+|NC)p(NC)} \\ &= \frac{\frac{90}{100} \times \frac{1}{100}}{\frac{90}{100} \times \frac{1}{100} + \frac{8}{100} \times \frac{99}{100}} \\ &\approx 0.1 \end{aligned}$$

Isn't it too low?

Example of Conditional Probability



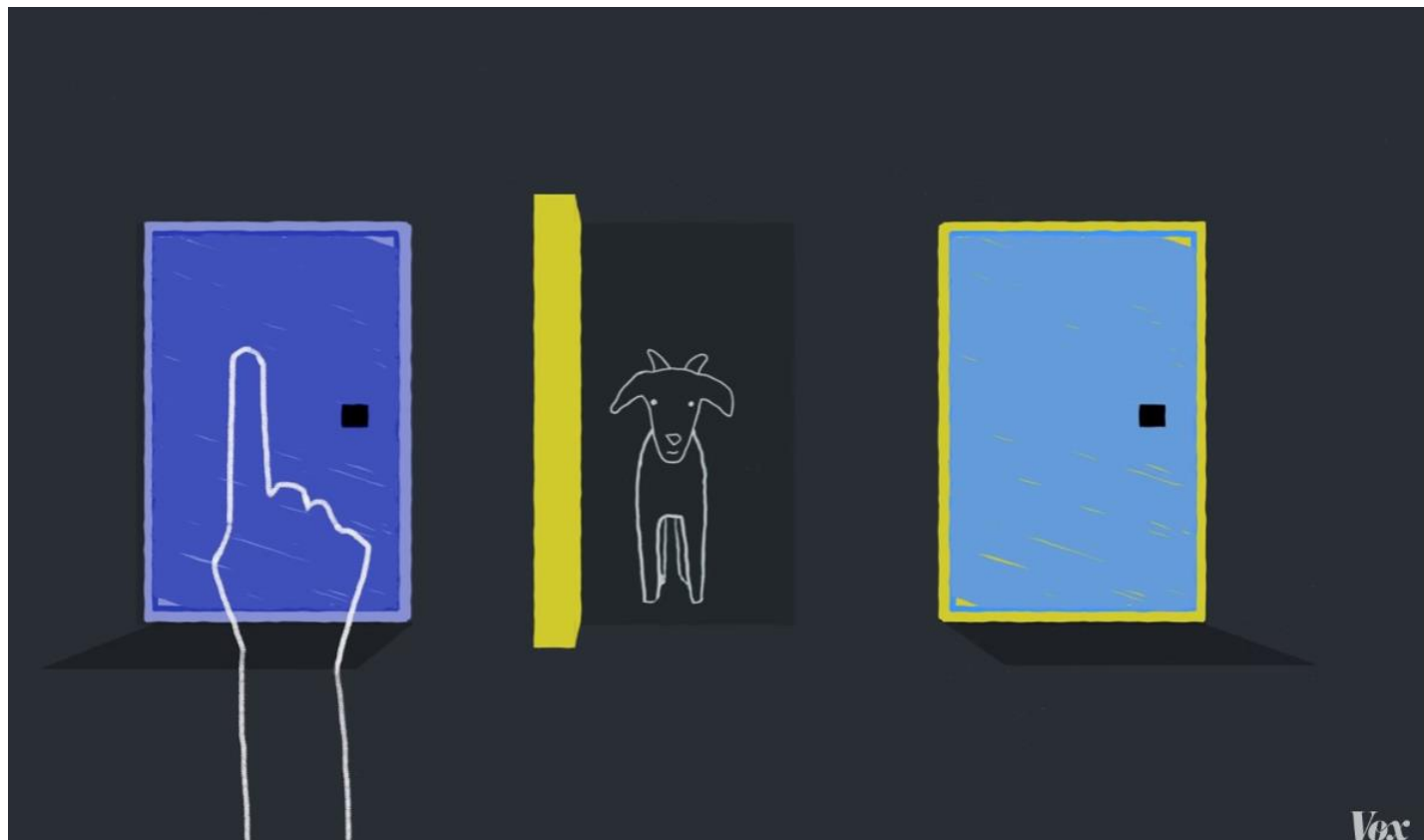
<https://www.youtube.com/watch?v=ggDQXlinbME>



Example of Conditional Probability



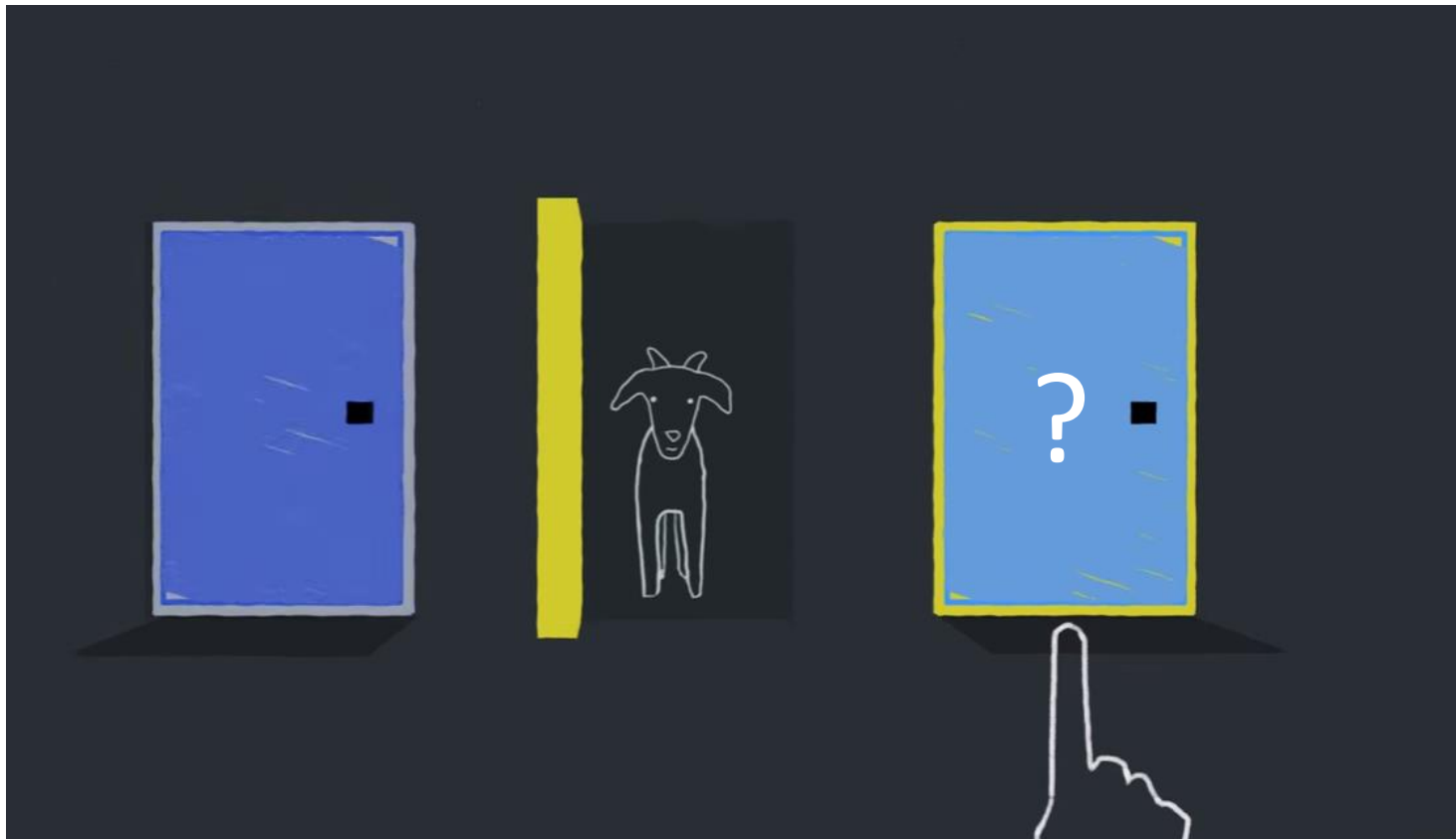
<https://www.youtube.com/watch?v=ggDQXlinbME>



Example of Conditional Probability



<https://www.youtube.com/watch?v=ggDQXlinbME>



Example of Conditional Probability

Assume we picked up Door 1 and then Monty shows us a goat behind Door 2

Event A: The car is behind Door 1 and

Event B: The Monty shows us a goat behind Door 2 (observation)

Example of Conditional Probability

Assume we picked up Door 1 and then Monty shows us a goat behind Door 2

Event A: The car is behind Door 1 and

Event B: The Monty shows us a goat behind Door 2

$$\begin{aligned}P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\&= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\sim A)P(\sim A)} \\&= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\text{Car bh D2})P(\text{Car bh D2}) + P(B|\text{Car bh D3})P(\text{Car bh D3})} \\&= \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{3}\end{aligned}$$

$\because P(B) = \sum_a P(B|A = a)P(A = a)$

$$P(\sim A|B) = \frac{2}{3} \quad \text{: Probability that the car is behind the door 3}$$

Changing the door will double the probability of getting the car!

Outline

1. Probability
- 2. Discrete distributions**
3. Continuous distributions

The Bernoulli Distribution

The probability distribution of a random variable which takes value 1 with success probability and value 0 with failure probability .

- **Notation**

$$Y \sim B(p)$$

$$p_Y(Y = y) = p(y) = B(y|p) = p^y(1 - p)^{1-y} \quad \text{for } y \in \{0,1\}$$

- **Parameters**

$p \in [0,1]$: Probability of success

- **Mean and variance**

$$E(Y) = p$$

$$\text{var}(Y) = p(1 - p)$$

The Binomial Distribution

The distribution on the number of successes in a sequence of n independent yes/no experiments

- The total number of heads among n coin tossing

- **Notation**

$$Y \sim \text{Bin}(n, p)$$

$$p(y) = \text{Bin}(y|n, p) = \binom{n}{y} p^y (1 - p)^{n-y} \quad y \in \{0, 1, \dots, n\}:$$

- **Parameters**

$n > 0$: Sample size (number of trials)

$p \in [0, 1]$: Probability of success

- **Mean and variance**

$$E(Y) = np$$

$$\text{var}(Y) = np(1 - p)$$

$n = 1$, the binomial distribution is a Bernoulli distribution

The Poisson Distribution

Expresses the probability of a given **number of events occurring in a fixed interval of time or space** if these events occur with a known average rate (λ)

- number of birth per hour on a given day in hospital
- disease cases within a given town

- **Notation**

$$Y \sim \text{Poisson}(\lambda)$$

$$P(y) = \text{Poisson}(y|\lambda) = \frac{1}{y!} \lambda^y \exp(-\lambda)$$

$$y \in \{0, 1, 2, \dots\},$$

- **Parameters**

$$\lambda > 0 : \text{rate}$$

- **Mean and variance**

$$E(Y) = \lambda$$

$$\text{var}(Y) = \lambda$$

The Poisson Distribution

$$\sum_y p(y|\lambda) = \sum_y \frac{\lambda^y e^{-\lambda}}{y!} = e^{-\lambda} \left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right] = e^{-\lambda} e^{\lambda} = 1$$

$$\begin{aligned} E[Y] &= \sum_y y \frac{\lambda^y e^{-\lambda}}{y!} = e^{-\lambda} \left[\lambda + \lambda^2 + \frac{\lambda^3}{2!} + \frac{\lambda^4}{3!} + \dots \right] = e^{-\lambda} \lambda \left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right] = \lambda \\ &= e^{-\lambda} \lambda e^{\lambda} = \lambda \end{aligned}$$

The Geometric Distribution

The probability distribution of the number X of Bernoulli trials needed to get one success

- **Notation**

$$Y \sim \text{Geometric}(\lambda)$$

$$p(y) = \text{Geometric}(y|\lambda) = (1 - \lambda)^{y-1} \lambda$$

$$y \in \{0, 1, 2, \dots\},$$

- **Parameters**

$\lambda > 0$: Probability of success

- **Mean and variance**

$$E(Y) = \frac{1}{\lambda}$$

$$\text{var}(Y) = \frac{1 - \lambda}{\lambda^2}$$

The Multinomial Distribution

the multinomial distribution is a generalization of the binomial distribution. the multinomial distribution gives the probability of any particular combination of numbers of successes for the various categories

- **Notation**

$$Y = (Y_1, \dots, Y_k) \sim \text{Multin}(n, p_1, \dots, p_k) \quad \textcolor{red}{y = (y_1, \dots, y_k) : \text{Vector!}}$$

$$p(y) = \text{Multin}(y|n, p_1, \dots, p_k) = \binom{n}{y_1 \ y_2 \ \dots \ y_k} p_1^{y_1} \dots p_k^{y_k}$$

$$y_j \in \{0, 1, \dots, n\}, \quad \sum_{j=1}^k y_j = n$$

- **Parameters**

$n > 0$: Sample size (number of trials)

$p_j \in [0, 1], \sum_{j=1}^k p_j = 1$: Probability of occurrence

- **Mean and variance**

$$E(Y_j) = np_j$$

$$\text{var}(Y_j) = np_j(1 - p_j)$$

$$\text{cov}(Y_i, Y_j) = -np_i p_j \text{ when } i \neq j$$

Outline

1. Probability
2. Discrete distributions
- 3. Continuous distributions**

The Uniform Distribution

- **Notation**

$$Y \sim U(\alpha, \beta)$$

$$p(y) = U(y|\alpha, \beta) = \frac{1}{\beta - \alpha}$$

$$y \in [\alpha, \beta]$$

- **Parameters**

α, β ($\beta > \alpha$) : Boundaries

- **Mean and variance**

$$E(Y) = \frac{\alpha + \beta}{2}$$

$$\text{var}(Y) = \frac{(\beta - \alpha)^2}{12}$$

The Univariate Normal Distribution

- **Notation**

$$Y \sim N(\mu, \sigma^2)$$

$$p(y) = N(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} (y - \mu)^2\right)$$
$$y \in [-\infty, \infty]$$

- **Parameters**

μ : Mean (location)

$\sigma > 0$: Standard deviation (scale)

- **Mean and variance**

$$E(Y) = \mu$$

$$\text{var}(Y) = \sigma^2$$

The Multivariate Normal Distribution

- **Notation**

$$Y = (Y_1, \dots, Y_k) \sim N(\mu, \Sigma)$$

$$p(y) = N(y|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2} (y - \mu)^T \Sigma^{-1} (y - \mu)\right)$$

$$y = (y_1, \dots, y_k) : \text{Vector! } y_i \in [-\infty, \infty]$$

- **Parameters**

$$\mu = (\mu_1, \dots, \mu_k) : \text{Mean vector}$$

$$\Sigma \succcurlyeq 0 : \text{Positive semi definite covariance matrix, } \Sigma_{i,j} = \text{cov}(y_i, y_j)$$

- **Mean and variance**

$$E(Y) = \mu$$

$$\text{var}(Y) = \Sigma$$

The Multivariate Normal Distribution

A random vector $X = (X_1, \dots, X_k)$ is a Gaussian random vector (GRV) if the joint pdf for X is of the form

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

Property 1 : For a GRV, uncorrelation implies independence

This can be verified by substituting $\sigma_{ij} = \text{Cov}(x_i, x_j) = 0$ for all $i \neq j$ in the joint pdf. Then Σ becomes diagonal and so does Σ^{-1} , and the joint pdf reduces to the product of the marginal $X_i \sim N(\mu_i, \sigma_{ii})$

The Multivariate Normal Distribution

Property 2 : Linear transformation of a GRV yields a GRV, i.e., given any $m \times n$ matrix A , where $m \leq n$ and A has full rank m , then

$$Z = AY \sim N(A\mu + b, A\Sigma A^T)$$

Because

$$E(Z) = E(AY + b) = AE(Y) + b = A\mu + b$$

$$\begin{aligned}\Sigma_Z &= E[(Z - E(Z))(Z - E(Z))^T] \\ &= E[(AY - A\mu)(AY - A\mu)^T] \\ &= AE[(Y - \mu)(Y - \mu)^T]A^T \\ &= A\Sigma A^T\end{aligned}$$

(The proof is not complete; we need to show the joint PDF is in a right form)

Example: Let

$$Y = N\left(\mathbf{0}, \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}\right)$$

Find the joint pdf of

$$Z = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} Y$$

The Multivariate Normal Distribution

proof of property 2

1. Assume we have a random vector $X = (X_1, \dots, X_k)$ is a Gaussian random vector (or X_1, \dots, X_k are jointly Gaussian r.v.s.) if the joint pdf is of the form

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^k |\det(\Sigma)|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

2. Assume $y = Ax$, $x = A^{-1}y$ with $A^{-1} = A^T(AA^T)^{-1}$

$Ax = AA^{-1}y = y$ with $A^{-1} = A^T(AA^T)^{-1}$ is right pseudo inverse
Check: $AA^T(AA^T)^{-1} = (AA^T)(AA^T)^{-1} = I$

3. we can compute the pdf for y using $p_X(x)$:

$$\begin{aligned} p_Y(y) &= \frac{1}{|\det(A)|} p_X(A^{-1}y) \\ &= \frac{1}{|\det(A)|} \frac{1}{\sqrt{(2\pi)^k |\det(\Sigma)|}} \exp\left[-\frac{1}{2}(A^{-1}y - \mu)^T \Sigma^{-1}(A^{-1}y - \mu)\right] \\ &= \frac{1}{\sqrt{(2\pi)^k |\det(A\Sigma A^T)|}} \exp\left[-\frac{1}{2}(y - A\mu)^T (A\Sigma A^T)^{-1}(y - A\mu)\right] \end{aligned}$$

$$\begin{aligned} \mu_Y &= A\mu \\ \Sigma_Y &= A\Sigma A^T \end{aligned}$$

The Multivariate Normal Distribution

Property 3 : Marginal of GRV are Gaussian, i.e., if Y is GRV then for any subset $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$ of indexes, the RV

$$Z = \begin{bmatrix} Y_{i_1} \\ Y_{i_2} \\ \vdots \\ Y_{i_k} \end{bmatrix} \text{ is GRV}$$

Converse is not generally true

To show this we use Property 2. For example, let $n = 3$ and $Z = \begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix}$, we can express Z as a linear transformation of Y :

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix}$$

Therefore,

$$Z = N \left(\begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix} \right)$$

The Multivariate Normal Distribution

Property 4 : Conditionals of a GRV are Gaussians, more specifically, if

$$Z = \begin{bmatrix} Y_1 \\ - \\ Y_2 \end{bmatrix} \sim N \left(\begin{bmatrix} Y_1 \\ - \\ Y_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & | & \Sigma_{12} \\ - & - & - \\ \Sigma_{21} & | & \Sigma_{22} \end{bmatrix} \right)$$

where Y_1 is k -dim RV and Y_2 is an $n - k$ dim RV, then

$$Y_2 | \{Y_1 = y\} \sim N(\Sigma_{21}\Sigma_{11}^{-1}(y - \mu_1) + \mu_2, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

Example:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim N \left(\begin{bmatrix} 1 \\ - \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & | & 2 & 1 \\ - & | & - & - \\ 2 & | & 5 & 2 \\ 1 & | & 2 & 9 \end{bmatrix} \right)$$

$$E \left(\begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix} \middle| Y_1 = y \right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} (y - 1) + \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2y \\ y + 1 \end{bmatrix}$$

$$\Sigma \left(\begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix} \middle| Y_1 = y \right) = \begin{bmatrix} 5 & 2 \\ 2 & 9 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$$

The Gamma Distribution

- The gamma distribution is frequently used to model waiting times.
- The arrival times in the Poisson process have gamma distributions

- **Notation**

$$Y \sim \text{Gamma}(\alpha, \beta)$$

$$p(y) = \text{Gamma}(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$
$$y > 0$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Gamma function $\Gamma(\alpha)$ is a normalization constant to ensure that the total probability integrates to 1.

- **Parameters**

$\alpha > 0$: Shape parameter

$\beta > 0$: Rate parameter (Inverse scale)

- **Mean and variance**

$$E(Y) = \frac{\alpha}{\beta}$$

$$\text{var}(Y) = \frac{\alpha}{\beta^2}$$

The Gamma Distribution

- The gamma distribution is frequently used to model waiting times.
- The arrival times in the Poisson process have gamma distributions

- **Notation**

$$Y \sim \text{Gamma}(\alpha, \beta)$$

$$p(y) = \text{Gamma}(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$
$$y > 0$$

- **Parameters**

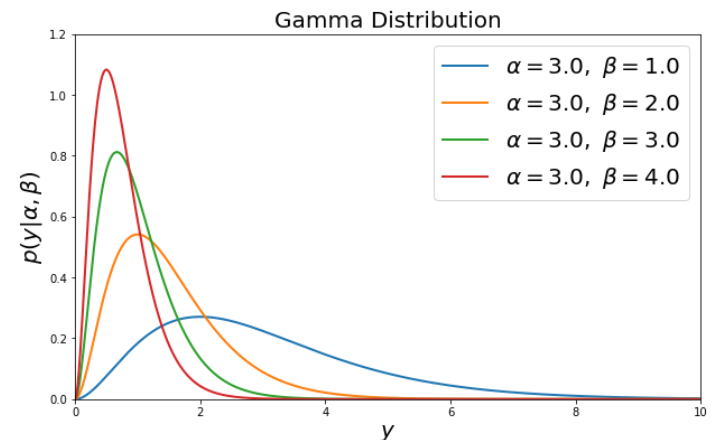
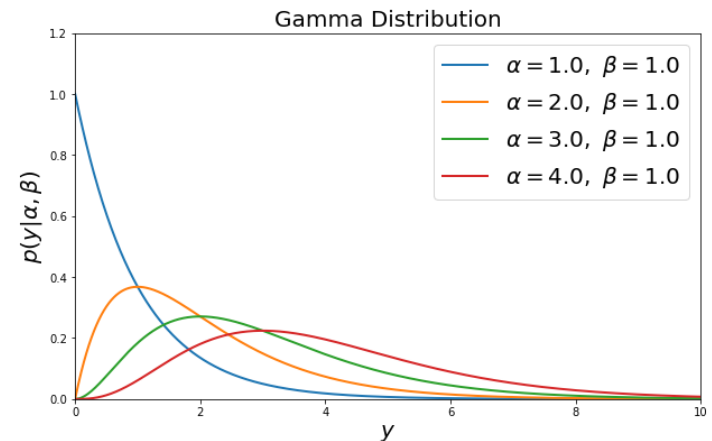
$\alpha > 0$: Shape parameters

$\beta > 0$: Inverse scale (rate) parameters

- **Mean and variance**

$$E(Y) = \frac{\alpha}{\beta}$$

$$\text{var}(Y) = \frac{\alpha}{\beta^2}$$



The Gamma Distribution

- The gamma distribution is frequently used to model waiting times.
- The arrival times in the Poisson process have gamma distributions

$$Y \sim \text{Gamma}(\alpha, \beta)$$

$$p(y) = \text{Gamma}(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$
$$y > 0$$

Properties

- $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$
- $\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}$ for $\lambda > 0$
- $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$
- $\Gamma(n) = (n - 1)!$, for $n = 1, 2, 3, \dots$

$$\begin{aligned} \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-\beta y} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{\beta^\alpha} = 1 \end{aligned}$$

The Gamma Distribution

$$\begin{aligned} E(Y) &= \int_0^{\infty} y \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha} e^{-\beta y} dy \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int_0^{\infty} \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} y^{\alpha} e^{-\beta y} dy \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\alpha \Gamma(\alpha)}{\beta^{\alpha+1}} \quad \because \Gamma(\alpha+1) = \alpha \Gamma(\alpha) \\ &= \frac{\alpha}{\beta} \end{aligned}$$

$$\text{Gamma}(y|\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$

$$\text{Gamma}(y|\alpha+1, \beta) = \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} y^{\alpha+1-1} e^{-\beta y}$$

The Exponential Distribution

- Model **the time between the occurrence of events** in an interval of time, or the distance between events in space.
 - ✓ The duration of a phone call to a help center
 - ✓ The time between successive failures of a machine
- The exponential random variable can be viewed as a continuous analogue of the geometric distribution

- Notation**

$$Y \sim \text{Exp}(\lambda)$$

$$p(y) = \text{Exp}(y|\lambda) = \begin{cases} \lambda e^{-\lambda y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

- Parameters**

$\lambda > 0$: the rate parameter

- Mean, variance and mode**

$$E(Y) = \frac{1}{\lambda}$$

$$\text{var}(Y) = \frac{1}{\lambda^2}$$

$$\text{Gamma}(1, \lambda) = \text{Exp}(\lambda)$$

$$\begin{aligned} \text{Gamma}(y|\alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \\ \text{Gamma}(y|1, \lambda) &= \frac{\lambda^1}{\Gamma(1)} y^{1-1} e^{-\lambda y} \\ \text{Gamma}(y|1, \lambda) &= \lambda e^{-\lambda y} \end{aligned}$$

The Beta Distribution

The beta distribution is a suitable model for the random behavior of percentages and proportions (i.e., a distribution *of probabilities*)

- **Notation**

$$Y \sim \text{Beta}(\alpha, \beta) \quad Y \in [0, 1]$$

$$p(y) = \text{Beta}(y|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1} (1-y)^{\beta-1}$$

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$

Beta function $B(\alpha, \beta)$ is a normalization constant to ensure that the total probability integrates to 1.

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

- **Parameters**

$\alpha, \beta > 0$: Shape parameters

- **Mean and variance**

$$E(Y) = \frac{\alpha}{\alpha + \beta}$$

$$\text{var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

The Inverse Gamma Distribution

- the distribution of the reciprocal of a variable distributed according to the gamma distribution
- Widely used for Bayesian statistics

- **Notation**

$$Y \sim \text{Inv} - \text{Gamma}(\alpha, \beta) \quad Y > 0$$

$$p(y) = \text{Inv} - \text{Gamma}(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} e^{-\beta/y}$$

- **Parameters**

$\alpha > 0$: Shape parameters

$\beta > 0$: Inverse scale (rate) parameters

- **Mean and variance**

$$E(Y) = \frac{\beta}{\alpha - 1} \quad \text{for } \alpha > 1$$

$$\text{var}(Y) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \quad \text{for } \alpha > 2$$

The Dirichlet Distribution

- Multivariate generalization of the Beta distribution
- Probability distribution of probability distribution

$$\theta \sim \text{Beta}(\alpha, \beta) \quad 0 \leq \theta \leq 1$$

$$Y \sim \text{Bin}(n, \theta)$$

$$\theta = (\theta_1, \dots, \theta_k) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k) \quad \sum_i^k \theta_i = 1$$

$$Y = (Y_1, \dots, Y_k) \sim \text{Multin}(n, \theta_1, \dots, \theta_k)$$

- **Notation**

$$Y = (Y_1, \dots, Y_k) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k)$$

$$p(y) = \text{Dirichlet}(y | \alpha_1, \dots, \alpha_k) = \frac{1}{\text{B}(\alpha)} y_1^{\alpha_1-1} \dots y_k^{\alpha_k-1}$$

$$= \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} y_1^{\alpha_1-1} \dots y_k^{\alpha_k-1}$$

$$\text{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

$$\text{B}(\alpha) = \frac{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \dots + \alpha_k)}$$

$$\text{B}(\alpha) = \frac{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)}{\Gamma(\alpha_0)}$$

$y = (y_1, \dots, y_k)$ is a probability simplex : $y_j \in [0,1]$, $\sum_j^k y_j = 1$

- **Parameters**

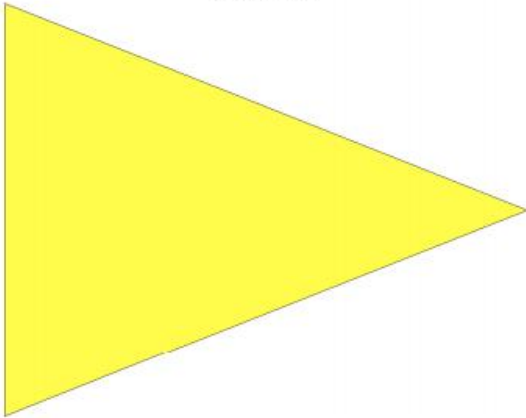
$$\alpha = (\alpha_1, \dots, \alpha_k), \alpha_i > 0, \quad \alpha_0 = \sum_{j=1}^k \alpha_j$$

- **Mean and variance**

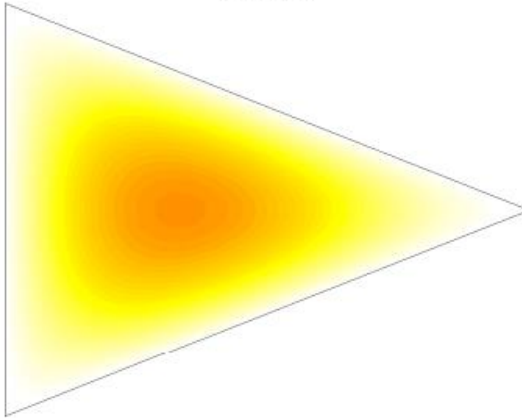
$$\text{E}(Y_j) = \frac{\alpha_j}{\alpha_0}, \quad \text{var}(Y_j) = \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0 + 1)} \quad \text{cov}(Y_i, Y_j) = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}, i \neq j$$

The Dirichlet Distribution

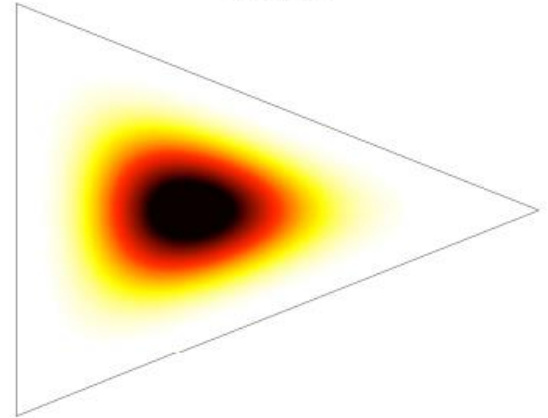
Dir(1.0,1.0,1.0)



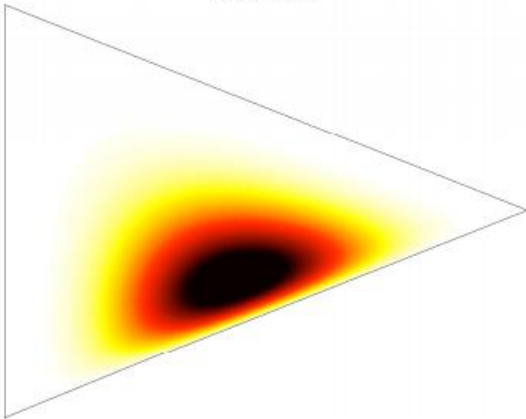
Dir(2.0,2.0,2.0)



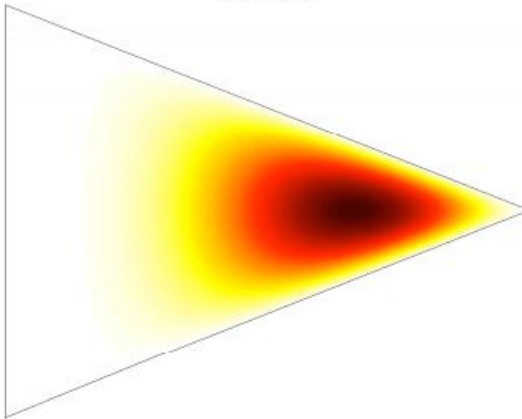
Dir(5.0,5.0,5.0)



Dir(5.0,5.0,2.0)



Dir(5.0,2.0,2.0)



Dir(0.7,0.7,0.7)

