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

$p_Y(Y = y) = p(y)$: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

$Y \sim \mathrm{dist}(\theta)$, θ is parameter for distribution
$p_Y(Y = y|\theta) = p(y|\theta) = \mathrm{dist}(y|\theta)$ or $p_Y(Y = y; \theta) = p(y; \theta) = \mathrm{dist}(y; \theta)$

Example :

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

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

Useful results from Probability Theory for Bayesian Statistics

$$p(u, v)$$

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

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

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

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

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

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

Mean and Variance of conditional distributions

$$\mathrm{E}(U) = \int up(u)du$$

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

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

Conditional mean (law of total expectation)

$$\mathrm{E}(U) = \mathrm{E}_V\big(\mathrm{E}(U|V)\big)$$

$$\because\ \mathrm{E}(U) = \int\int up(u,v)dudv = \int\int up(u|v)du\ p(v)dv = \int E(U|v)p(v)dv = \mathrm{E}_V\big(E(U|V)\big)$$

Conditional variance (law of total variance)

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

Proof:

$$
\begin{aligned}
\mathrm{var}(U) &= \mathrm{E}_U[U^2] - (\mathrm{E}_U[U])^2 \\
&= \mathrm{E}_V\left[\mathrm{E}_U[U^2|V]\right] - \left\{\mathrm{E}_V\left[\mathrm{E}_U[U|V]\right]\right\}^2 && \because \text{The law of total expectation} \\
&= \mathrm{E}_V[\mathrm{var}(U|V) + (\mathrm{E}_U[U|V])^2] - \left\{\mathrm{E}_V\left[\mathrm{E}_U[U|V]\right]\right\}^2 && \because \mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \\
&= \mathrm{E}_V[\mathrm{var}(U|V)] + \mathrm{E}_V[(\mathrm{E}_U[U|V])^2] - \left\{\mathrm{E}_V\left[\mathrm{E}_U[U|V]\right]\right\}^2 && \because \text{Linearity of expectation} \\
&= \mathrm{E}_V[\mathrm{var}(U|V)] + \mathrm{E}_V[(\mathrm{E}_U[U|V])^2] - \left\{\mathrm{E}_V\left[\mathrm{E}_U[U|V]\right]\right\}^2 \\
&= \mathrm{E}_V[\mathrm{var}(U|V)] + \mathrm{var}(\mathrm{E}_U[U|V])
\end{aligned}
$$

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

$$p(C|+) = \frac{p(C,+)}{p(+)} = \frac{p(+|C)p(C)}{p(+)}$$

$$= \frac{p(+|C)p(C)}{p(+|C)p(C)\ + p(+|NC)p(NC)}$$

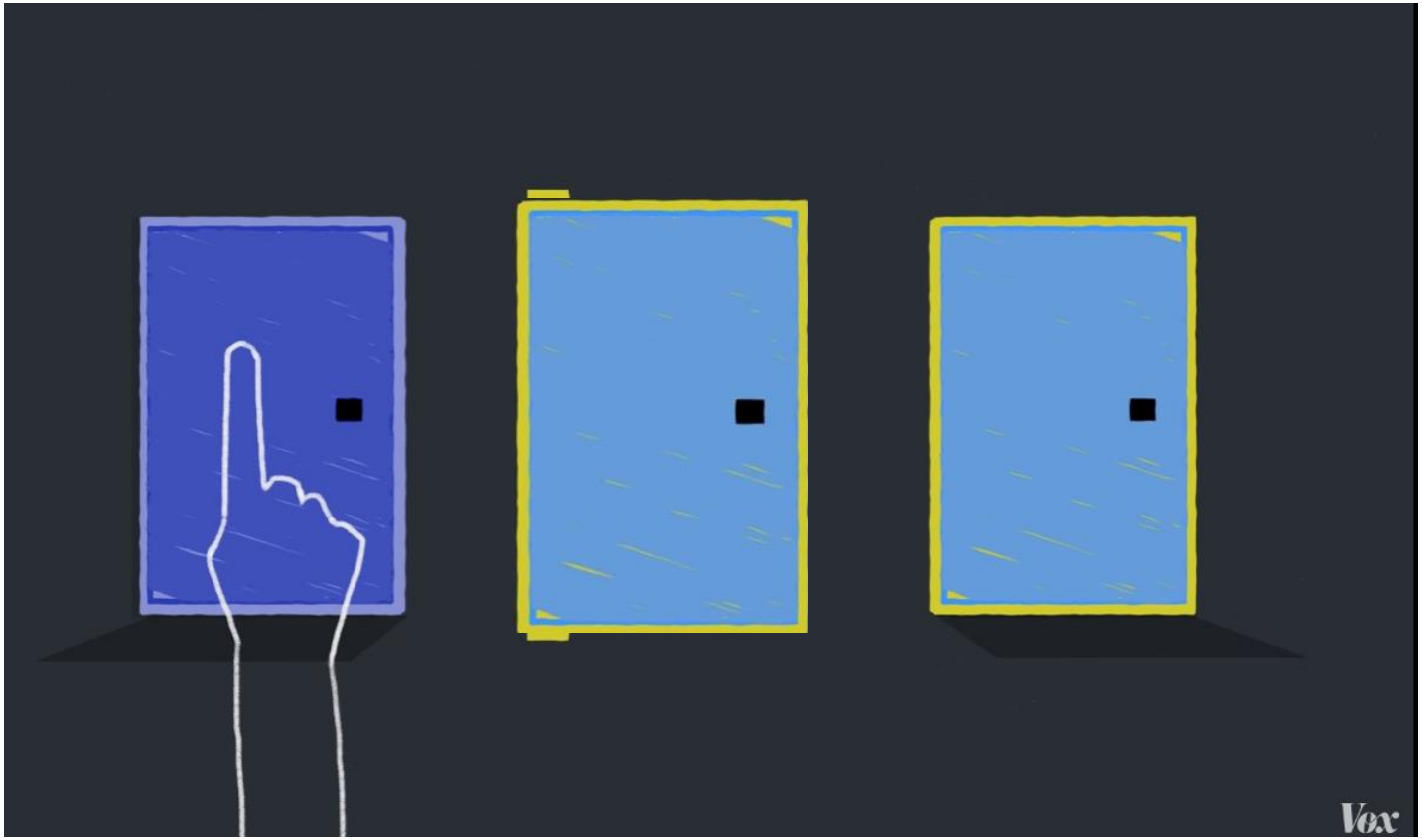
$$= \frac{\frac{90}{100} \times \frac{1}{100}}{\frac{90}{100} \times \frac{1}{100} + \frac{8}{100} \times \frac{99}{100}}$$

$$\approx 0.1$$

Isn't it too low?

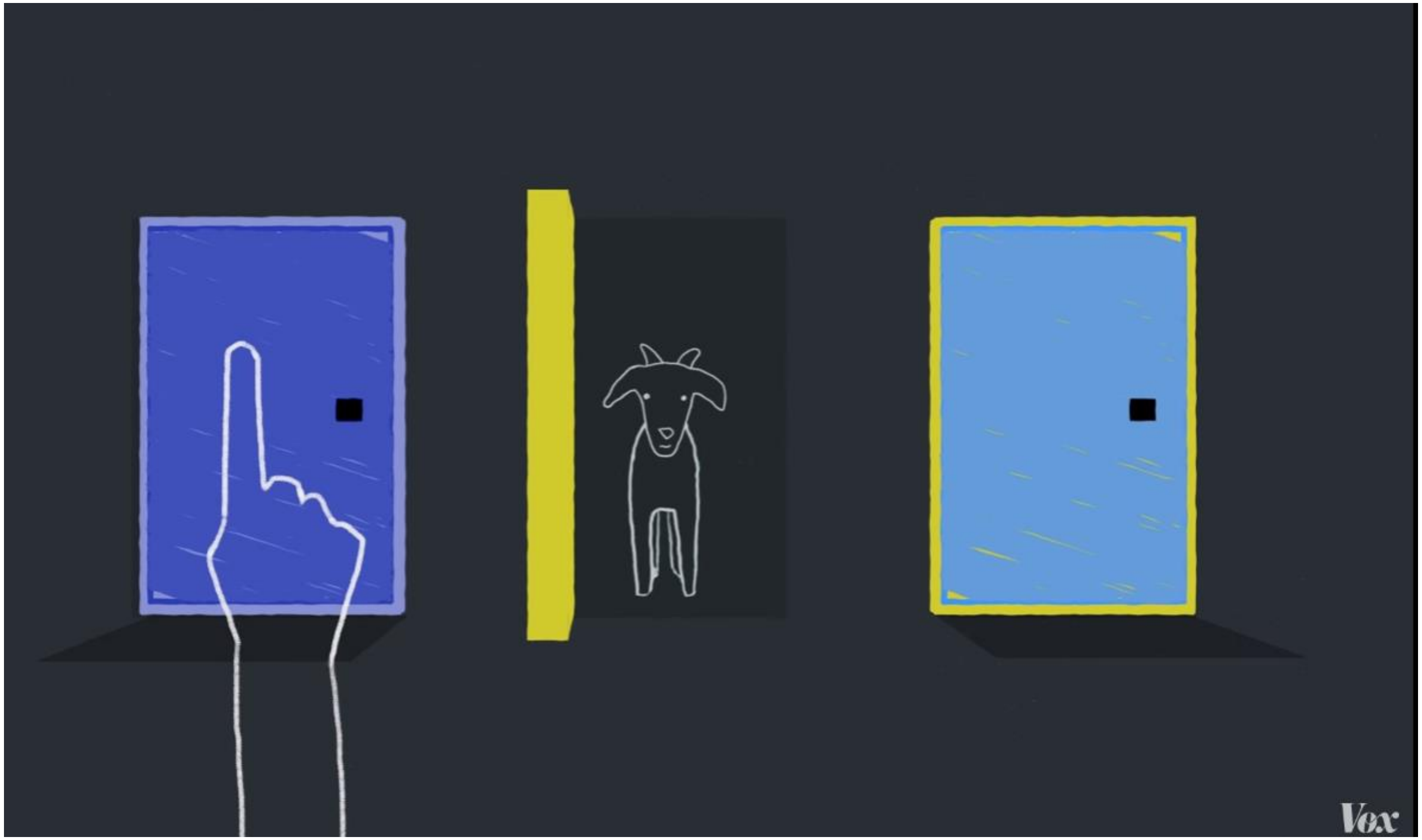
Example of Conditional Probability

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



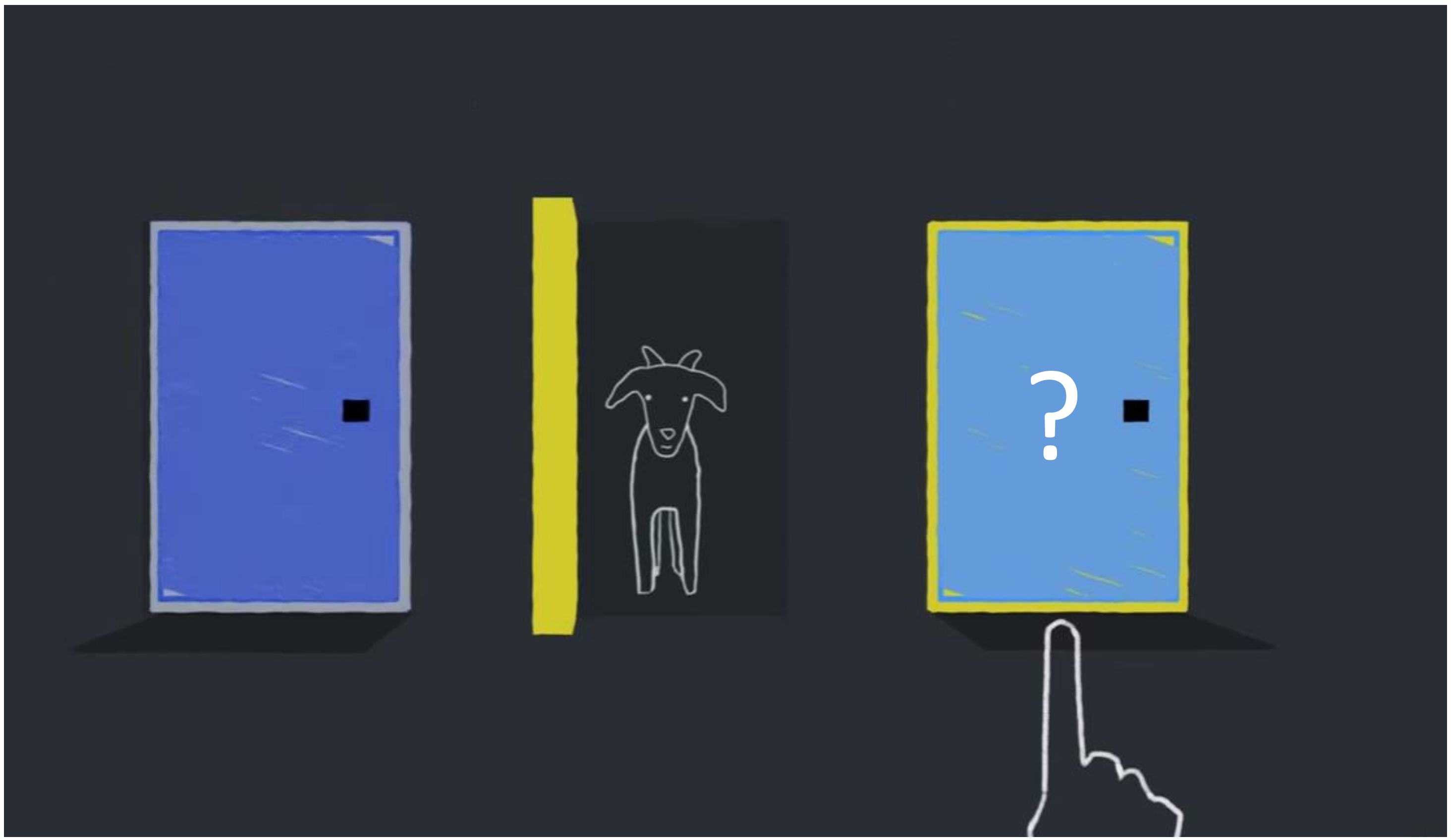
Example of Conditional Probability

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

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

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

$$P(\sim A|B) = \frac{2}{3}$$: 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**

$Y \sim \mathrm{B}(p)$

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

- **Parameters**

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

- **Mean and variance**

$\mathrm{E}(Y) = p$

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

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**

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

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

- **Parameters**

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

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

- **Mean and variance**

$$\mathrm{E}(Y) = np$$

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

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 (lambda)

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

- **Notation**

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

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

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

- **Parameters**

$\lambda > 0$: rate

- **Mean and variance**

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

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

The Poisson Distribution

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

$$\mathrm{E}[Y] = \sum_y y\frac{\lambda^y e^{-\lambda}}{y!} = e^{-\lambda}\left[\lambda + \lambda^2 + \frac{\lambda^3}{2!} + \frac{\lambda^4}{3!} + \cdots\right] = e^{-\lambda}\lambda\left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right] = \lambda$$

$$= e^{-\lambda}\lambda e^{\lambda} = \lambda$$

The Geometric Distribution

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

- **Notation**

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

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

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

- **Parameters**

$\lambda > 0$: Probability of success

- **Mean and variance**

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

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

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**

$$Y = (Y_1, \dots, Y_k) \sim \mathrm{Multin}(n, p_1, \dots, p_k)$$

$y = (y_1, \dots, y_k)$: Vector!

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

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

- **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**

$\mathrm{E}(Y_j) = np_j$

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

$\mathrm{cov}(Y_i, Y_j) = -np_ip_j$ when $i \neq j$

Outline

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

The Uniform Distribution

- **Notation**

$$Y \sim \mathrm{U}(\alpha, \beta)$$

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

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

- **Parameters**

α, β $(\beta > \alpha)$: Boundaries

- **Mean and variance**

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

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

The Univariate Normal Distribution

- **Notation**

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

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

$$y \in [-\infty, \infty]$$

- **Parameters**

μ : Mean (location)

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

- **Mean and variance**

$$\mathrm{E}(Y) = \mu$$

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

The Multivariate Normal Distribution

- **Notation**

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

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

$y = (y_1, \ldots, y_k)$: Vector! $y_i \in [-\infty, \infty]$

- **Parameters**

$\mu = (\mu_1, \ldots, \mu_k)$: Mean vector

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

- **Mean and variance**

$\mathrm{E}(Y) = \mu$

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

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 $\sigma_{ij} = Cov\left(x_i, x_j\right) = 0$ 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 $X_i \sim N(\mu_i, \sigma_{ii})$

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

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

Because

$$E(Z) = \mathrm{E}(AY) = A\mathrm{E}(\mathrm{Y}) = \mathrm{A}\mu$$

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

(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

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

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 To compute the probability of the transformed random vector y utilizing the existing pdf $p_X(x)$, there should be one-to-one relationship between X and Y

→ $y = Ax$ have a one solution only when A is full column rank matrix (A ∈ $R^{m\times n}$, with $rank = n$)

→ solution $x = (A^T A)^{-1} A^T y$

3. Assuming the left inverse $(A^T A)^{-1} A^T$ is equivalent to A^{-1}, we can compute the pdf for y using $p_X(x)$:

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

$$\mu_Y = A\mu$$
$$\Sigma_Y = A\Sigma A^T$$

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 $\mathrm{Z} = \begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix}$, 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,

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

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

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

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:

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

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

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

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**

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

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

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**

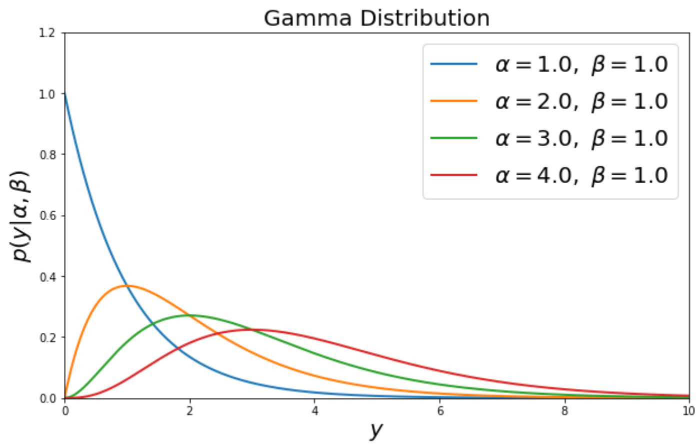
$\alpha > 0$: Shape parameters

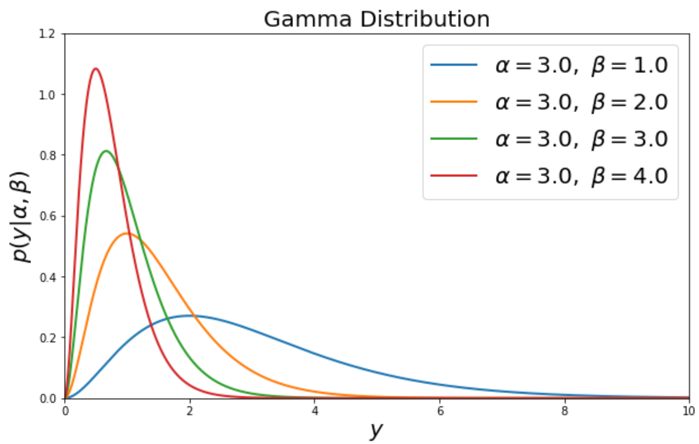
$\beta > 0$: Inverse scale (rate) parameters

- **Mean and variance**

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

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





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 $n = 1,2,3, \ldots$

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

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{\beta^\alpha} = 1$$

The Gamma Distribution

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

$\because \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

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

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

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**

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

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

- **Parameters**

$\lambda > 0$: the rate parameter

- **Mean, variance and mode**

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

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

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

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

$$\mathrm{Gamma}(y|1, \lambda) = \frac{\lambda^1}{\Gamma(1)} y^{1-1} e^{-\lambda y}$$

$$\mathrm{Gamma}(y|1, \lambda) = \lambda e^{-\lambda y}$$

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**

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

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

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

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

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

- **Parameters**

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

- **Mean and variance**

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

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

The Inverse Gamma Distribution

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

- **Notation**

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

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

- **Parameters**

$\alpha > 0$: Shape parameters

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

- **Mean and variance**

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

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

The Dirichlet Distribution

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

$$\theta \sim \mathrm{Beta}(\alpha, \beta) \quad 0 \le \theta \le 1 \qquad \theta = (\theta_1, \dots, \theta_k) \sim \mathrm{Dirichlet}(\alpha_1, \dots, \alpha_k) \quad \Sigma_i^k \theta_i = 1$$

$$Y \sim \mathrm{Bin}(n, \theta) \qquad Y = (Y_1, \dots, Y_k) \sim \mathrm{Multin}(n, \theta_1, \dots, \theta_k)$$

- **Notation**

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

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

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

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

$$\mathrm{B}(\alpha) = \frac{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \cdots, \alpha_k)}$$

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

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

- **Parameters**

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

- **Mean and variance**

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

The Dirichlet Distribution