VE401 Recitation 3

Discrete Random Variable

Hypergeometric Distribution

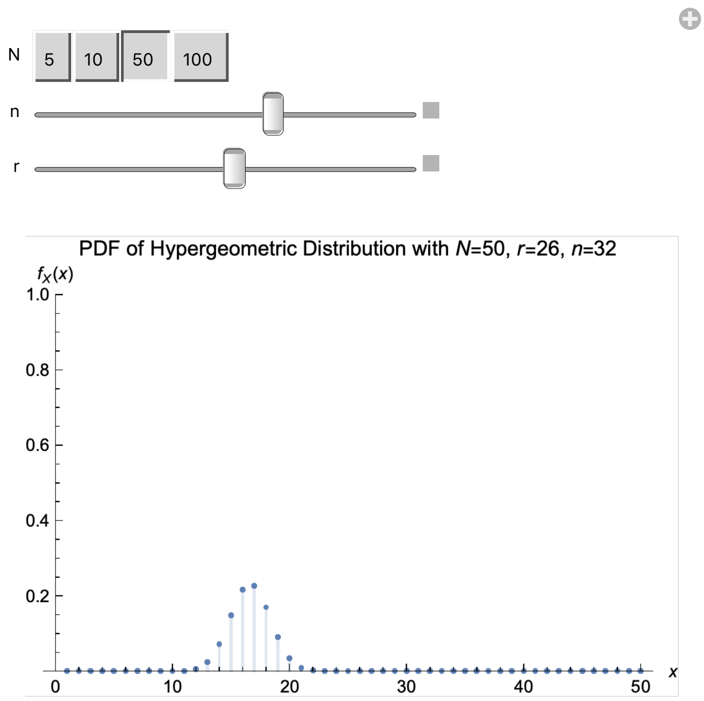
Purpose: Select n samples from N objects (within which r objects have trait), what is the probability of having x objects with trait?

Parameter and properties:

- $N \in \{0, 1, 2, \ldots\}$ is the number of total objects.
- $n \in \{0, 1, 2, \ldots, N\}$ is the number of samples.
- $r \in \{0, 1, 2, \ldots, N\}$ is the number of objects with traits. $p := \frac{r}{N}$, $q := 1 - p$.

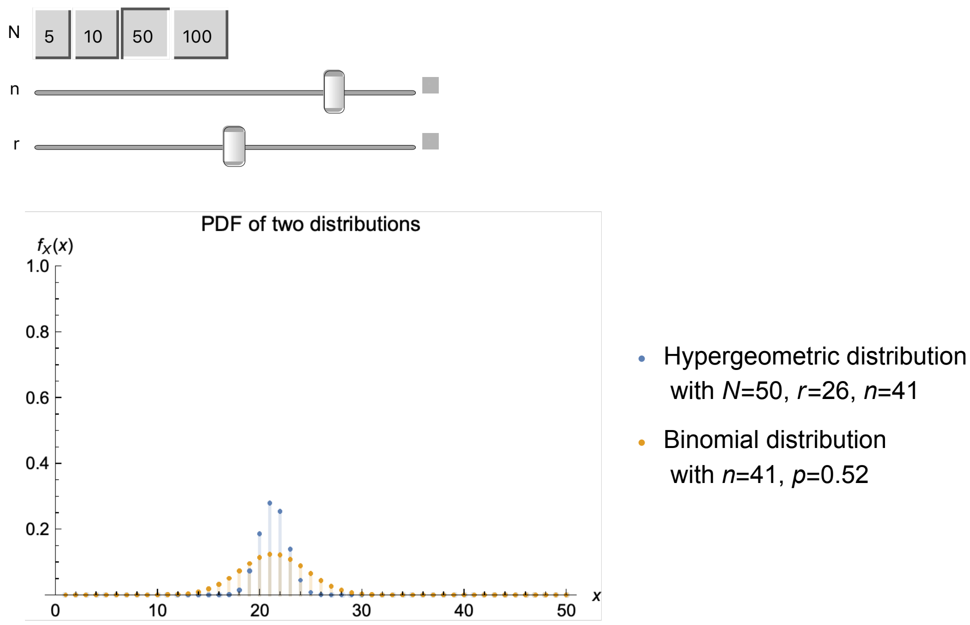
$\mathrm{E}[X]$	$\mathrm{Var}[X]$	PDF
$n\,p$	$n\,p\,q\,\frac{N-n}{N-1}$	$\begin{cases} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} & 0 \le x \le N \\ 0 & \text{otherwise} \end{cases}$

Out[ ]=



Approximation: if n / N is sufficiently small, it can be approximated by binomial distribution with parameter n and $p = r / N$.

Out[∘]=



Example:

I have a batch of 1000 products, within which 5 are defective. My friend said that he will not accept the batch if there are more than 1% of products that are defective. If he is collecting 100 samples, what is the probability of my batch being falsely rejected?

$$1 - P[X \leq 1] = 1 - \sum_{x=0}^{1} \frac{\binom{5}{x}\binom{995}{100-x}}{\binom{1000}{100}} = 1 - 91.898\,\% = 8.102\,\%$$

Or we can use the binomial approximation,

$$1 - P[X \leq 1] = 1 - \sum_{x=0}^{1} \binom{100}{x} (0.005)^x\,(0.995)^{100-x} = 1 - 91.018\,\% = 8.982\,\%$$

Summary

Discrete Random Variables

(a) Distribution

	Geometric	Binomial	Pascal	Hypergeometric	Poisson
Ω	$\mathbb{N}\setminus\{0\}$	$\{0,1,2,\dots,n\}$	$\{r,r+1,\dots\}$	Ω_H	$\mathbb{N}$
P	$0<p<1$ $q=1-p$	$0<p<1$ $q=1-p$ $n\in\mathbb{N}\setminus\{0\}$	$0<p<1$ $q=1-p$ $r\in\mathbb{N}\setminus\{0\}$	$N,n,r\in\mathbb{N}\setminus\{0\}$ $r,n\le N$	$k=\lambda t$ $k\in\mathbb{R}$
f_X	$(1-p)^{x-1}p$	$\binom{n}{x}p^x(1-p)^{n-x}$	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\frac{k^x e^{-k}}{x!}$
F	$1-q^{\lfloor x\rfloor}$				
m_X	$(-\infty,-\ln q)\to\mathbb{R}$ $\frac{pe^t}{1-qe^t}$	$\mathbb{R}\to\mathbb{R}$ $(q+pe^t)^n$	$\mathbb{R}\to\mathbb{R}$ $\frac{(pe^t)^r}{(1-qe^t)^r}$		$\mathbb{R}\to\mathbb{R}$ $e^{k(e^t-1)}$
E	$\frac{1}{p}$	np	$\frac{r}{p}$	$n\frac{r}{N}$	k
Var	$\frac{q}{p^2}$	npq	$\frac{rq}{p^2}$	$n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$	k

Note: $\Omega_H=\{x\in\mathbb{N}: max(0,n-(N-r))\le x\le min(n,r)\}$

(b) Approximation

Hypergeometric: binomial with n and $p=\frac{r}{N}$ when $\frac{n}{N}$ is sufficiently small.

Binomial: poisson with $k=pn$ when n is large and p is small.

Continuous Random Variables

Chi-squared Distribution

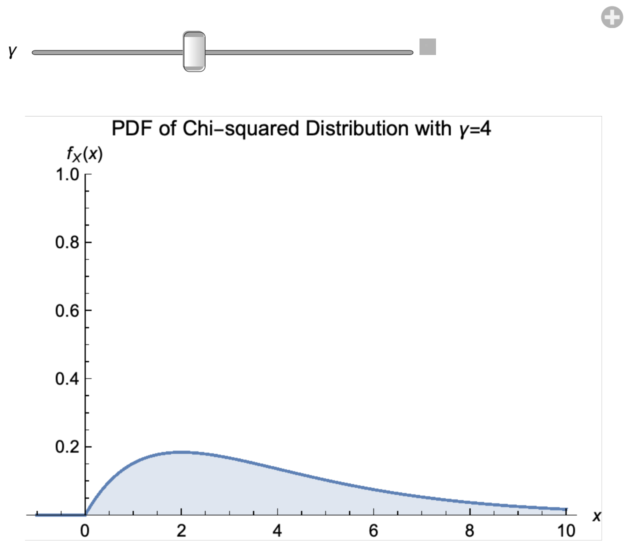
Purpose: how is the sum of squares of γ independent standard normal random variables distributed?

Parameter and properties:

- $\gamma\in\mathbb{N}^+$ is the ***degrees of freedom***.

Mean	Variance	PDF	MGF
γ	γ^2	$\begin{cases}\frac{1}{\Gamma\left(\frac{\gamma}{2}\right)2^{\gamma/2}}x^{\frac{\gamma}{2}-1}e^{-x/2} & x>0\\ 0 & \text{True}\end{cases}$	$(1-2\,t)^{-\gamma/2}$

Out[ ]=



Chebyshev Inequality

Purpose: To (roughly) estimate the variance of a random variable.

Mathematical representation: for $k \in \mathbb{N}\setminus\{0\}$, $P[\,|X|\, \geq c] \leq \frac{\mathrm{E}[|X|^k]}{c^k}$.

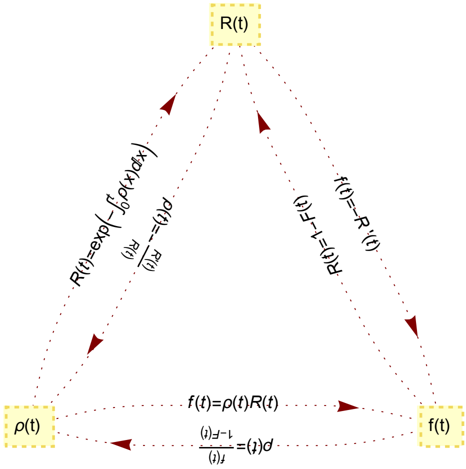
Application: $P[\,|X-\mu|\, \geq m\,\sigma] \leq \frac{1}{m^2}$.

Reliability

To study the reliability of a system, we have the following:

- ***Failure density*** $f(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t+\Delta t]}{\Delta t}$, the probability of failing at time t, (uniform, Weibull, ...)
- ***reliability function*** $R(t) = 1 - F(t)$, the probability of system still working at time t,
- ***hazard rate*** $\rho(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t+\Delta t | T \geq t]}{\Delta t} = \frac{f(t)}{R(t)}$, the rate of failing at time t **given that it didn't fail before t**.

Out[159]=



- For series system, all components still working ⇒ system still working.
  $R_S(t) = P[\text{all components still working}] = \prod_{i=1}^{k} R_i(t)$.
- For parallel system, at least one component still working ⇒ system still working.
  $R_S(t) = P[\text{not all components broken}] = 1 - \prod_{i=1}^{k} (1 - R_i(t))$.

Example:

Your professor has announced that we will have a quiz on Monday class (just an example! not real!), but you don't know the exact time. What is the probability of not having a quiz in the first 45 minutes? Assume that a class is non-stop and 90 minutes long.

We can use uniform distribution to describe the "failure" density,

$$f(t) = \begin{cases} \frac{1}{90} & \text{if } 0 \le t \le 90 \\ 0 & \text{otherwise} \end{cases}$$

And the reliability at time $t = 45$ is

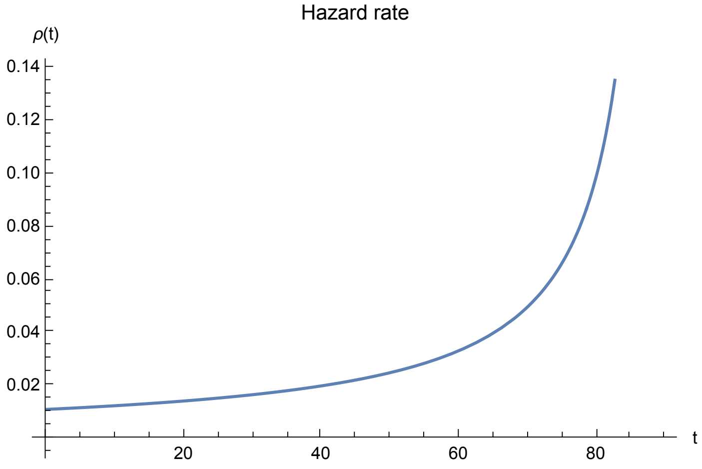
$$R(t) = 1 - F(t) = 1 - t/90 = 0.5$$

what is the expected rate of your professor giving the quiz at the very next moment?

The hazard rate at $t = 45$ can be calculated as

$$\rho(t) = \frac{f(t)}{1-F(t)} = \frac{1/90}{1-t/90} = \frac{1/90}{1-45/90} = \frac{1}{45}$$

Out[144]=



Weibull Distribution

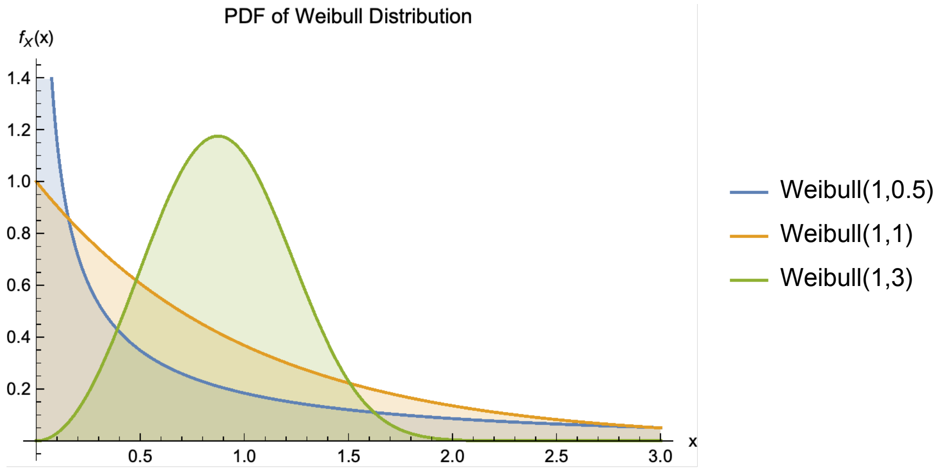
Purpose: To represent a failure density $f(t)$.

Parameters and properties:

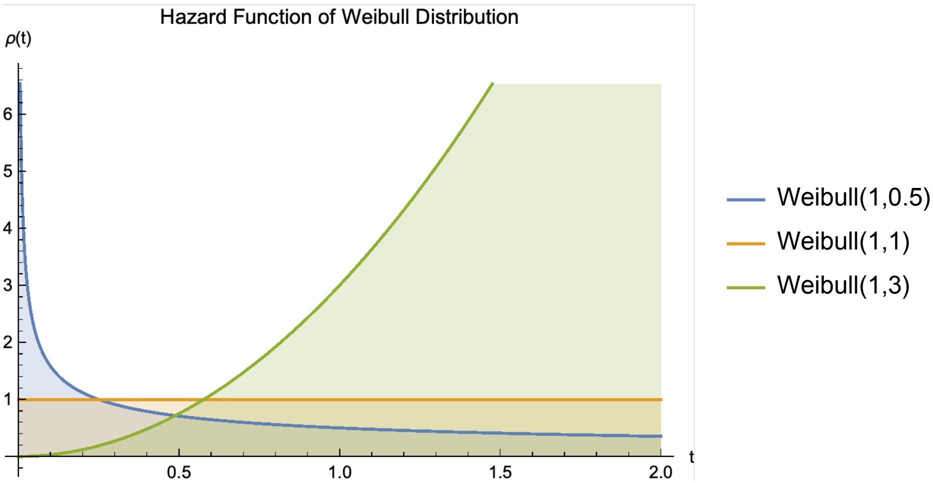
- $\alpha,\ \beta > 0$.
- the hazard rate is $\begin{cases} \text{constant} & \text{if } \beta = 1 \\ \text{increasing} & \text{if } \beta > 1 \\ \text{decreasing} & \text{if } \beta < 1 \end{cases}$.

Mean	Variance	PDF
$\alpha^{-1/\beta}\,\Gamma\left(1+\frac{1}{\beta}\right)$	$\alpha^{-2/\beta}\left(\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma\left(1+\frac{1}{\beta}\right)^2\right)$	$\begin{cases} \alpha\beta\, x^{\beta-1}\, e^{-\alpha\, x^{\beta}} & x>0 \\ 0 & \text{True} \end{cases}$

Out[226]=



Out[229]=



Summary

Distribution

	P	f_X	m_X	E	Var
Exponential	$\beta \in \mathbb{R}$ $\beta > 0$	$\begin{cases} \frac{1}{\beta}e^{-x/\beta}, & x > 0, \\ 0, & x \le 0, \end{cases}$	$(-\infty, \frac{1}{\beta}) \to \mathbb{R}$ $(1-\beta t)^{-1}$	β	β^2
Gamma	$\alpha, \beta \in \mathbb{R}$ $\alpha, \beta > 0$	$\begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}, & x > 0, \\ 0, & x \le 0, \end{cases}$	$(-\infty, \frac{1}{\beta}) \to \mathbb{R}$ $(1-\beta t)^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
Normal	$\mu \in \mathbb{R}$ $\sigma > 0$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-((x-\mu)/\sigma)^2/2}$	$\mathbb{R} \to \mathbb{R}$ $e^{\mu t+\sigma^2 t^2/2}$	μ	σ^2
Standard Normal	$\mu = 0$ $\sigma = 1$	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$	$\mathbb{R} \to \mathbb{R}$ $e^{t^2/2}$	0	1
Weibull (two-parameter)	$\alpha, \beta > 0$	$\begin{cases} \alpha\beta x^{\beta-1}e^{-\alpha x^\beta}, & x > 0, \\ 0, & x \le 0, \end{cases}$		μ_W	σ_W^2
Uniform		$\begin{cases} 1, & 0 \le x \le 1, \\ 0, & otherwise \end{cases}$			

Note:

- $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1}e^{-z}, \alpha > 0$: Euler gamma function.
- $\Gamma(1) = 1, \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ if $\alpha > 1$.
- $\mu_W = \alpha^{-1/\beta}\Gamma(1+1/\beta), \sigma_W^2 = \alpha^{-2/\beta}\Gamma(1+2/\beta) - \mu^2$.

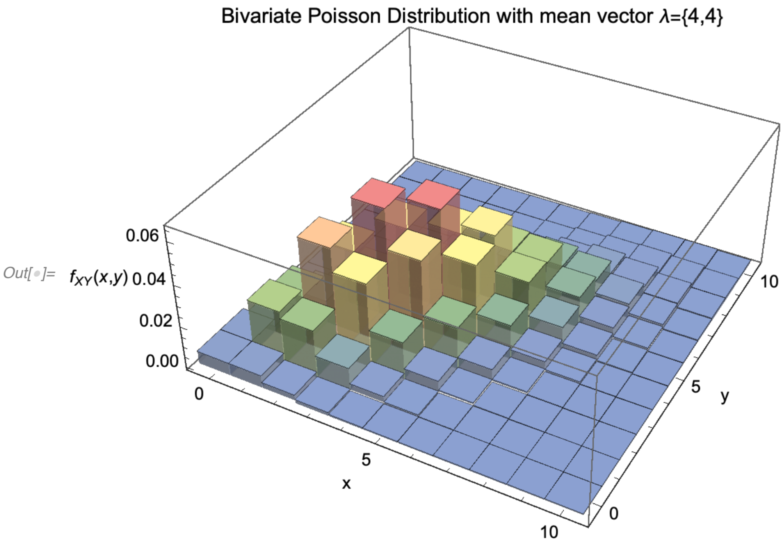
Joint Distributions

Discrete Multivariate Random Variables and its Properties

A ***discrete multivariate random variable*** is a map $\boldsymbol{X} : S \to \Omega$ with a PDF $f_{\boldsymbol{X}} : \Omega \to \mathbb{R}$ such that

- $f_{\boldsymbol{X}}(x, y) \ge 0$ for all $x = (x_1, \ldots, x_n) \in \Omega$,
- $\sum_{x \in \Omega} f_{\boldsymbol{X}}(x) = 1$.

where $f_{\boldsymbol{X}}(x) = P[\boldsymbol{X} = x = (x_1, \ldots, x_n)]$. The ***marginal density*** for X is $f_X(x) = \sum_y f_{XY}(x, y)$.



Example:

Let X, Y be random variables such that their PDF looks like

$f_{XY}(x, y)$	$y = -1$	$y = 0$	$y = 1$	$f_X(x)$
$x = 0$	0	$\frac{1}{3}$	0	
$x = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$	
$f_Y(y)$				

, calculate the marginal density $f_X(x)$ and $f_Y(y)$.

The marginal density is easily calculated as

$f_{XY}(x, y)$	$y = -1$	$y = 0$	$y = 1$	$\boldsymbol{f_X(x)}$
$x = 0$	0	$\frac{1}{3}$	0	$\boldsymbol{\frac{1}{3}}$
$x = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\boldsymbol{\frac{2}{3}}$
$\boldsymbol{f_Y(y)}$	$\boldsymbol{\frac{1}{3}}$	$\boldsymbol{\frac{1}{3}}$	$\boldsymbol{\frac{1}{3}}$	$\boldsymbol{1}$

Continuous Multivariate Random Variables and its Properties

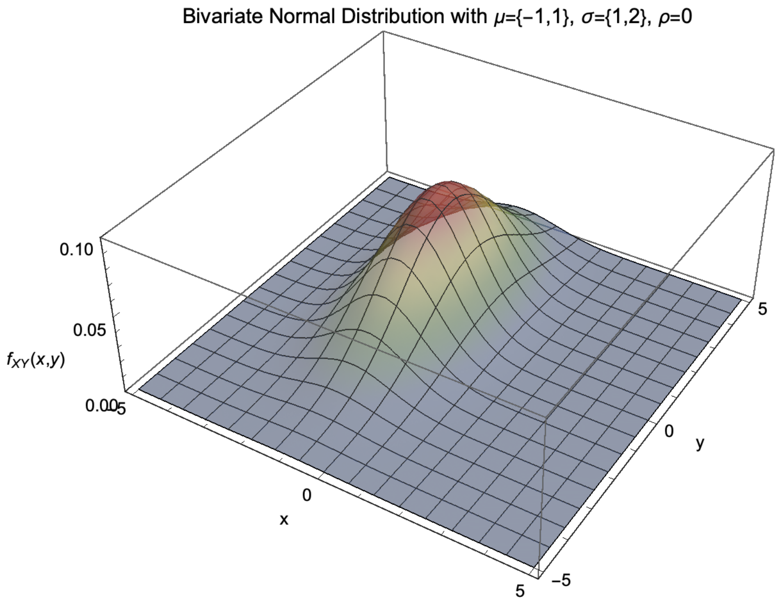
A ***continuous multivariate random variable*** is a map $\boldsymbol{X} : S \to \mathbb{R}^2$ with a PDF $f_{XY} : \mathbb{R}^2 \to \mathbb{R}$ such that

- $f_{\boldsymbol{X}}(x) \geq 0$ for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,
- $\int_{\mathbb{R}^n} f_{\boldsymbol{X}}(x)\, dx = 1$.

where $P[\boldsymbol{X} \in \Omega] = \int_{\Omega} f_{\boldsymbol{X}}(x)\, dx$.

The ***marginal density*** for X is $f_{X_k}(x_k) = \int_{\mathbb{R}^{n-1}} f_{\boldsymbol{X}}(x)\, dx_1 \ldots dx_{k-1}\, dx_{k+1} \ldots dx_n$.

Out[ ]=



Independence and Conditional Densities

Comparing two RVs is very analogous to comparing two events.

	Independence	Conditional Probability
two events A and B	$P[A \cap B] = P[A]\, P[B]$	$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$
two RVs X and Y	$f_{XY}(x,\ y) = f_X(x)\, f_Y(y)$ for all $(x,\ y) \in \operatorname{dom} f_{XY}$	$f_{X\mid y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}$

Where $f_{X\mid y}(x) = P[X = x \mid Y = y]$.

Example:

Let $X,\ Y$ be random variables such that their PDF looks like

$f_{XY}(x,\ y)$	$y = -1$	$y = 0$	$y = 1$	$f_X(x)$
$x = 0$	0	$\frac{1}{3}$	0	$\frac{1}{3}$
$x = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
$f_Y(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

, calculate $P[X = 0 \mid Y = 1]$.

This is given by $P[X = 0 \mid Y = 1] = \frac{P[X=0 \text{ and } Y=1]}{P[Y=1]} = \frac{0}{1/3} = 0$.

Expectation

Again the properties of expectation of bivariate RVs are similar to RV with single variable.

	Discrete Case	Continuous Case
$\mathrm{E}[\varphi \circ \boldsymbol{X}]$	$\sum_{x \in \Omega} \varphi(x)\, f_X(x)$	$\int_{\mathbb{R}^n} \varphi(x)\, f_X(x)\, dx$

$E[X_k]$	$E[X_k] = \sum_{x\in\Omega} x_k\, f_{\mathbf{X}}(x)$	$E[X_k] = \int_{\mathbb{R}^n} x_k\, f_{\mathbf{X}}(x)\, dx$
$E[X_i + X_j]$	$E[X_i + X_j] = E[X_i] + E[X_j]$	
$E[X_i \mid x_j]$	$E[X_i \mid X_j = x_j] = \sum_{x_i} x_i\, f_{X_i\mid x_j}(x_j)$	$E[X_i \mid X_j = x_j] = \int_{-\infty}^{\infty} x_i\, f_{X_i\mid x_j}(x_j)\, dx_j$

Example:

Let X, Y be random variables such that their PDF looks like

$f_{XY}(x,\ y)$	$y = -1$	$y = 0$	$y = 1$	$f_X(x)$
$x = 0$	0	$\frac{1}{3}$	0	$\frac{1}{3}$
$x = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
$f_Y(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

, calculate $E[X]$, $E[Y \mid X = 1]$ and $E[X\, Y]$.

We have $E[X] = 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = \frac{2}{3}$, $E[Y \mid X = 1] = -1\cdot\frac{1/3}{2/3} + 0\cdot\frac{0}{2/3} + 1\cdot\frac{1/3}{2/3} = 0$ and $E[X\, Y] = -1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = 0$.

Covariance

Intuition: The tendency in the linear relationship. Does more X lead to more Y?

Mathematical representation: Simple calculation on the variance of the sum of RV will give

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\, E\left[\left(X - E[X]\right)\left(Y - E[Y]\right)\right]$$

The extra term will be called the covariance of $(X,\ Y)$,

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

We note that

- $\mathrm{Cov}[X,\ X] = \mathrm{Var}[X]$,
- $\mathrm{Cov}[X,\ Y] = E[X\, Y] - E[X]\, E[Y] = 0$ if X and Y are independent. However, the converse is not true. i.e. zero covariance does not imply independence.

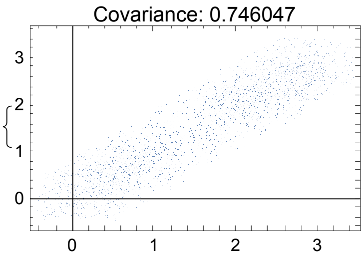
Example:

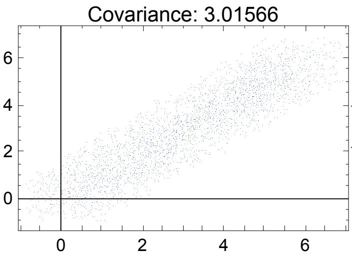
Let X, Y be random variables, and $\mathrm{Cov}[X,\ Y] = a$. What will $\mathrm{Cov}[2\, X,\ 2\, Y]$ be?

Simple calculation shows that

$$\mathrm{Cov}[2\, X,\ 2\, Y] = E[4\, X\, Y] - E[2\, X]\, E[2\, Y] = 4\, E[X\, Y] - 4\, E[X]\, E[Y] = 4\, a$$

So the covariance is dependent on the scale of data.

Out[36]= {


,


}

Let X, Y be random variables such that their PDF looks like

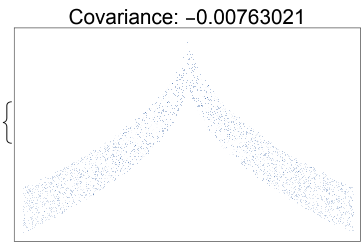
$f_{XY}(x, y)$	$y = -1$	$y = 0$	$y = 1$	$f_X(x)$
$x = 0$	0	$\frac{1}{3}$	0	$\frac{1}{3}$
$x = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
$f_Y(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

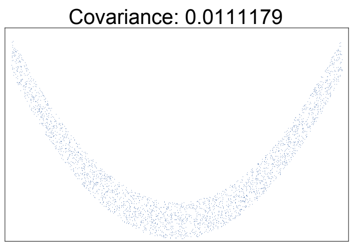
,

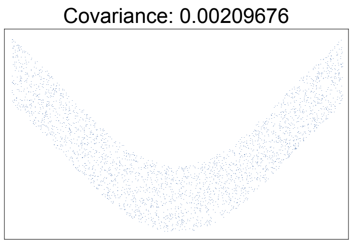
are they independent? Calculate Cov[X, Y].

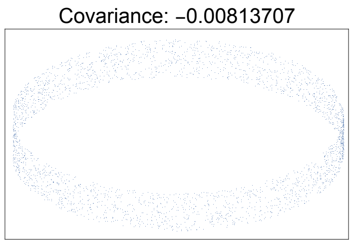
Since $P[x = 0 \text{ and } y = -1] = 0$ instead of $1/9$, X and Y are clearly not independent. However, we have $\mu_X = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$, $\mu_Y = \frac{1}{3} \cdot 0 = 0$, and

$$\begin{aligned} \mathrm{Cov}[X, Y] &= \mathrm{E}[X\,Y] - \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= 0 - 0 = 0 \end{aligned}$$

Out[ ]= {


,


,


,


}

Covariance Matrix

We define the ***covariance matrix*** for a multivariate RV,

$$\mathrm{Var}[\boldsymbol{X}] = \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_1, X_2] & \mathrm{Var}[X_2] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{Cov}[X_{n-1}, X_n] \\ \mathrm{Cov}[X_1, X_n] & \cdots & \mathrm{Cov}[X_{n-1}, X_n] & \mathrm{Var}[X_n] \end{pmatrix}$$

Properties:

- Covariance matrix is symmetric.
- For any constant $n \times n$ matrix with real coefficient $C \in \mathrm{Mat}(n \times n; \mathbb{R})$, $\mathrm{Var}[C\,\boldsymbol{X}] = C\,\mathrm{Var}[\boldsymbol{X}]\,C^T$.

Example:

Let X, Y be random variables such that their PDF looks like

$f_{XY}(x, y)$	$y = -1$	$y = 0$	$y = 1$	$f_X(x)$
$x = 0$	0	$\frac{1}{3}$	0	$\frac{1}{3}$
$x = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
$f_Y(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

, calculate the covariance matrix $\mathrm{Var}[(X, Y)]$.

We can first calculate

$$\mathrm{Var}[X] = \mathrm{E}\left[(X-\mu_X)^2\right] = \frac{1}{3}\left(0-\frac{2}{3}\right)^2 + \frac{2}{3}\left(1-\frac{2}{3}\right)^2 = \frac{2}{9}$$

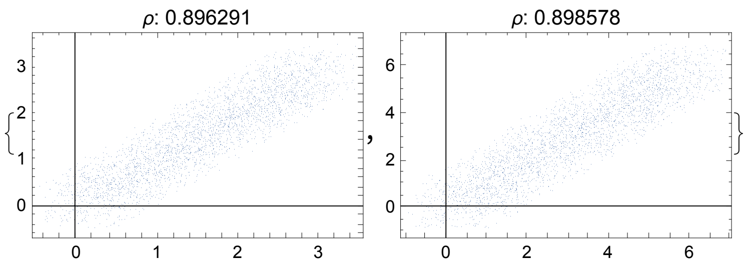
$$\mathrm{Var}[Y] = \mathrm{E}\left[(Y-\mu_Y)^2\right] = \frac{1}{3}(-1-0)^2 + \frac{1}{3}(0-0)^2 + \frac{1}{3}(1-0)^2 = \frac{2}{3}$$

Therefore we have $\mathrm{Var}[(X, Y)] = \begin{pmatrix} \mathrm{Var}[X] & \mathrm{Cov}[X, Y] \\ \mathrm{Cov}[X, Y] & \mathrm{Var}[Y] \end{pmatrix} = \begin{pmatrix} \frac{2}{9} & 0 \\ 0 & \frac{2}{3} \end{pmatrix}$.

Pearson Correlation Coefficient

Purpose: to eliminate the effect of the scale of data on covariance.

Mathematical representation: $\rho_{XY} = \dfrac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}$.

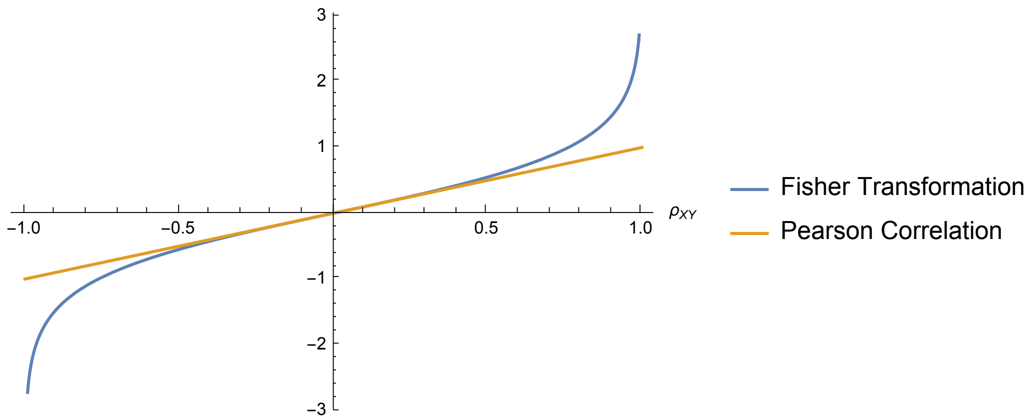
Out[44]=


Properties:

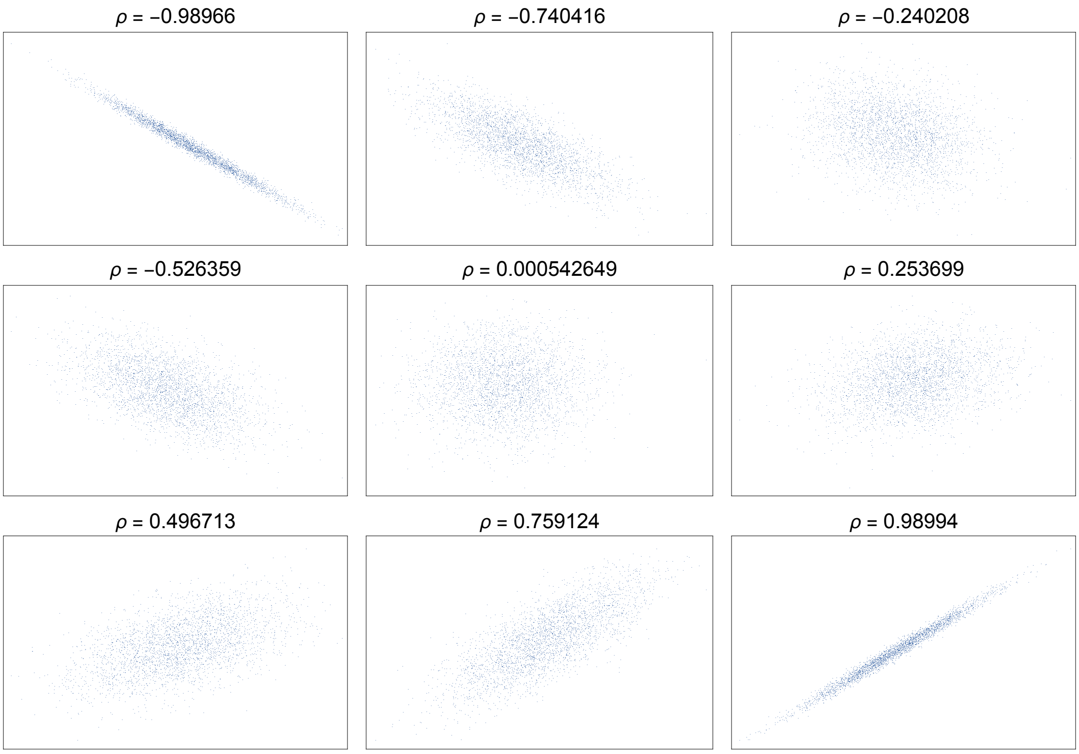
- $-1 \leq \rho_{XY} \leq 1$,
- $|\rho_{XY}| = 1$ indicates perfect linear relationship.
- To measure linearity of X and Y, the ***Fisher transformation*** $\frac{1}{2}\ln\left(\frac{1+\rho_{XY}}{1-\rho_{XY}}\right) = \mathrm{artanh}(\rho_{XY})$ is used.

  If $|\mathrm{artanh}(\rho_{XY})| \to \infty$ or $|\rho_{XY}| \to 1$, then it indicates linear relationship of X and Y.

Out[240]=



Out[84]=



Bivariate Normal Distribution

Purpose: Combine two normal distributed RVs together.

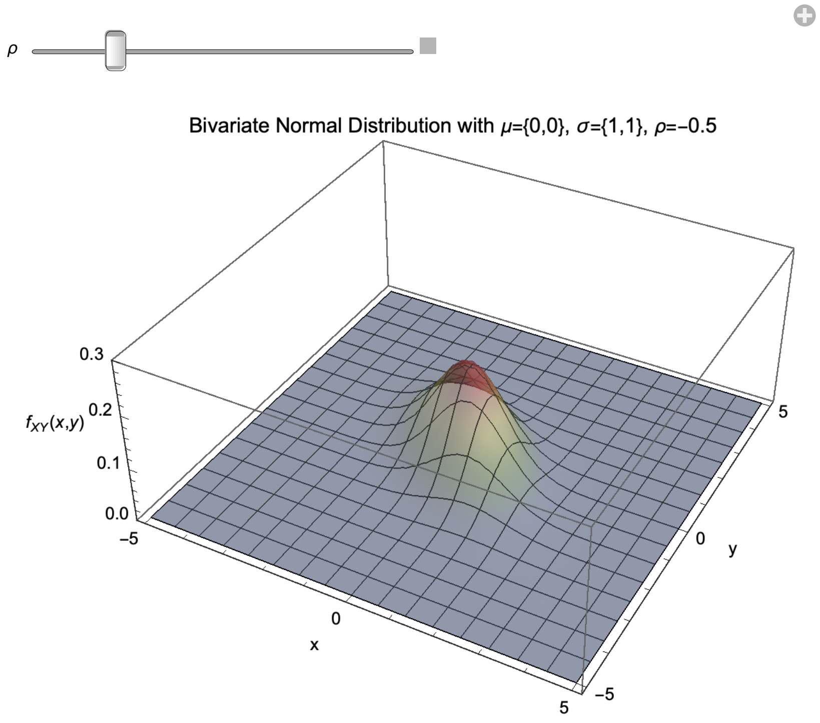
Parameter and properties:

- μ_X, μ_Y are the expectation of X and Y, respectively.
- σ_X, σ_Y are the standard deviation of X and Y, respectively.
- $-1 < \varrho < 1$ is the covariance of X and Y.

Mean	Variance	PDF
$\mathrm{E}[X] = \mu_X$ $\mathrm{E}[Y] = \mu_Y$	$\mathrm{Var}[X] = \sigma_X$ $\mathrm{Var}[Y] = \sigma_Y$	$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}}\exp\left(-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$

- The conditional mean of Y is linearly dependent on x, $\mathrm{E}[Y \mid x] = \mu_Y + \varrho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$. Taking $\mu_X = \mu_Y = 0$ and $\sigma_X = \sigma_Y = 1$ we have $\mathrm{E}[Y \mid x] = \varrho\, x$.

Out[250]=



Transformation of Variables

Let $((X, Y), f_{XY})$ be continuous bivariate random variable and $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ be differentiable bijective map with inverse φ^{-1}, then $(U, V) = \varphi \circ (X, Y)$ is a continuous bivariate random variable with

$$f_{UV}(u, v) = f_{XY} \circ \varphi^{-1}(u, v) \cdot \left| \det D\varphi^{-1}(u, v) \right|$$

where $D\varphi^{-1}(u, v) = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$ is the Jacobian of φ^{-1}.

Problems for Discussion

Transformation of Variables: Sum of Two Independent Random Variables

Problem: Given two independent random variables X, Y, and their PDF $f_X(x)$, $f_Y(y)$, how is $Z = X + Y$ distributed? We will start by investigating easier examples.

Discrete Case: Rolling Dice

Let X and Y denote the outcome of rolling a die. That is,

$$f_X(x) = f_Y(x) = \begin{cases} \frac{1}{6} & \text{if } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

Calculate $P[Z = 2]$, $P[Z = 3]$ and $P[Z = 4]$.

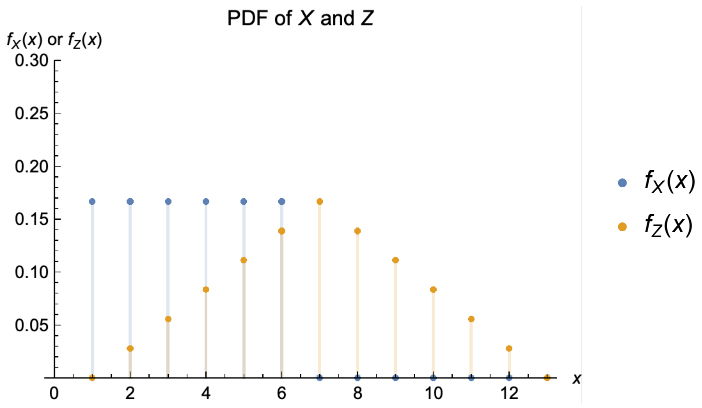
Solution

$$P[Z = 2] = P[X = 1]\ P[Y = 1] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

$$P[Z = 3] = P[X = 1]\, P[Y = 2] + P[X = 2]\, P[Y = 1] = \frac{2}{36}$$

And similarly, $P[Z = 4] = P[X = 1]\, P[Y = 3] + P[X = 2]\, P[Y = 2] + P[X = 3]\, P[Y = 1] = \frac{3}{36}$.

Out[236]=



In fact, you have proved in assignment 2 that

$$P[Z = z] = \sum_{x+y=z} P[X = x]\, P[Y = y] = \sum_{y=-\infty}^{\infty} f_X(z - y)\, f_Y(y)$$

We notice that this is actually the ***discrete convolution*** of f_X and f_Y.

Continuous Case: Sum of Uniform Distributed RVs

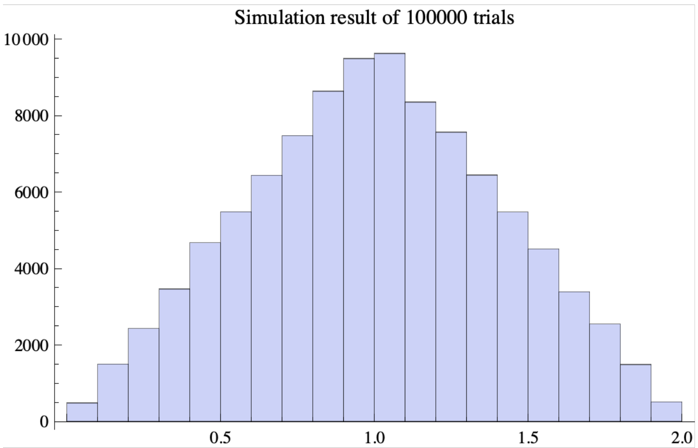
Let X and Y be identical independent uniform distributed random variables, such that

$$f_X(x) = f_Y(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the distribution of Z?

Solution

Out[190]=



It looks like a triangular PDF. Let us verify this. We define $\begin{pmatrix} u \\ v \end{pmatrix} = \varphi(x, y) := \begin{pmatrix} x+y \\ y \end{pmatrix}$, then $\begin{pmatrix} x \\ y \end{pmatrix} = \varphi^{-1}(u, v) = \begin{pmatrix} u-v \\ v \end{pmatrix}$. Therefore, the Jacobian of φ^{-1} is

$$D\varphi^{-1}(u, v) = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial u-v}{\partial u} & \frac{\partial u-v}{\partial v} \\ \frac{\partial v}{\partial u} & \frac{\partial v}{\partial v} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

and its determinant is 1. As a result,

$$\begin{aligned} f_{UV}(u, v) &= f_{XY} \circ \varphi^{-1}(u, v) \cdot \left|\det D\varphi^{-1}(u, v)\right| \\ &= f_{XY}(u-v, v) \\ &= f_X(u-v)\, f_Y(v) \end{aligned}$$

Calculating marginal density and replacing u with z,

$$f_Z(z) = f_U(z) = \int_{-\infty}^{\infty} f_X(z-v)\, f_Y(v)\, dv = (f_X * f_Y)(z)$$

So we can conclude that **f_Z is actually the *convolution* of f_X and f_Y.** From this we can easily calculate

$$f_Z(z) = \begin{cases} z & \text{if } 0 \le z \le 1 \\ 2-z & \text{if } 1 < z \le 2 \\ 0 & \text{otherwise} \end{cases}$$

which matches our experiment.

Application: Moment Generating Functions of Sum of Independent RVs

If X and Y are independent and moment generating function for X and Y are $m_X(t)$ and $m_Y(t)$, respectively, calculate $m_Z(t)$.

Solution

$$
\begin{aligned}
m_Z(t) &= \mathrm{E}\left[e^{t\,Z}\right] \\
&= \int_{-\infty}^{\infty} e^{z\,t}\, f_Z(z)\, d z \\
&= \int_{-\infty}^{\infty} e^{z\,t} \int_{-\infty}^{\infty} f_X(z-v)\, f_Y(v)\, d v\, d z \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{(z-v)\,t}\, f_X(z-v)\, e^{v\,t}\, f_Y(v)\, d v\, d z \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x\,t}\, f_X(x)\, e^{y\,t}\, f_Y(y)\, d y\, d x \qquad (x := z - v,\ y := v) \\
&= \int_{-\infty}^{\infty} e^{x\,t}\, f_X(x)\, d x \cdot \int_{-\infty}^{\infty} e^{y\,t}\, f_Y(y)\, d y \\
&= m_X(t)\, m_Y(t)
\end{aligned}
$$

Application: Sum of Two i.i.d. Exponential Random Variable

Let $f_X(x) = f_Y(x) = \begin{cases} b\, e^{-b\,x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$ be two exponential distributed PDF. What is the distribution of $Z = X + Y$?

Solution

Using the convolution property we can calculate that when $z \geq 0$,

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-v)\, f_Y(v)\, d v \\
&= \int_0^z b\, e^{-b\,(z-v)}\, b\, e^{-b\,v}\, d v \\
&= b^2\, z\, e^{-b\,z}
\end{aligned}
$$

Recall that gamma distribution has PDF $f_U(u) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, u^{\alpha-1}\, e^{-\beta\,u}$. We can see that $f_Z(z)$ is actually a PDF of gamma distribution with $\alpha = 2$ and $\beta = b$.

Another Solution

We have

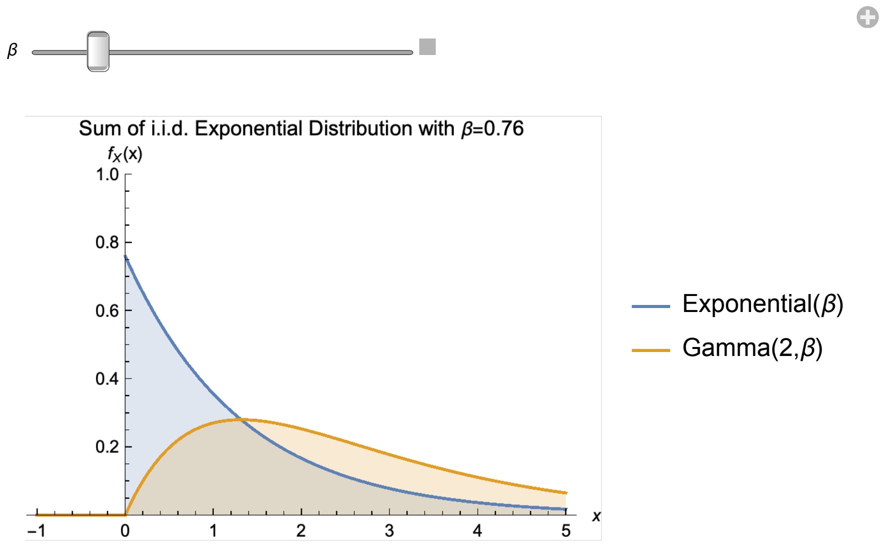
$$
\begin{aligned}
m_Z(t) &= m_X(t)\, m_Y(t) \\
&= \left(\frac{b}{b-t}\right)^2 \\
&= \left(1 - \frac{1}{b}\, t\right)^{-2}
\end{aligned}
$$

But gamma distribution with parameters α and β has MGF $\left(1 - \frac{t}{\beta}\right)^{-\alpha}$. Together with the **uniqueness of moment generating function**, we can conclude that Z follows a gamma distribution with $\alpha = 2$, $\beta = b$. This matches our intuition:

$$\text{time until second arrival} = \text{time until first arrival} + \text{time until first arrival}$$

given that the rate of arrival remains the same.

Out[135]=

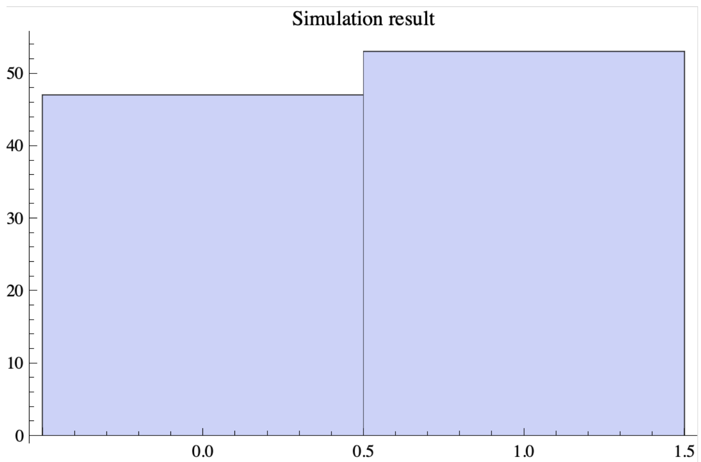


Application of Chebyshev Inequality: Weak Law of Large Numbers

Problem: We want to prove that when the number of independent and identical experiment is very large (e.g. flipping a coin 10000 times), then the proportion of trials with A (e.g. heads coming up) occurring will be close to actual probability of A happening (e.g. 0.5),

In[251]:=
```
Histogram[RandomInteger[{0, 1}, 100],
 PlotTheme → "Classic", PlotLabel → "Simulation result"]
```

Out[251]=



$$P\left[\left|\frac{\text{number of times } A \text{ occurs}}{\text{number of times experiment is performed}} - P[A]\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0 \quad \text{for any } \varepsilon > 0.$$

We can denote X_1, ..., X_n as **n i.i.d. Bernoulli random variable with mean $\mu = p$ and variance $\sigma^2 = p\,q$** to represent n experiments, where $P[X_i = 1] = P[A \text{ occuring on } i^{\text{th}} \text{ experiment}]$, then

$$\frac{\text{number of times } A \text{ occurs}}{\text{number of times experiment is performed}} = \frac{X_1 + \ldots + X_n}{n}$$

and $P[A] = \mu = p$.

Proof

Let $X := \frac{\text{number of times } A \text{ occurs}}{\text{number of times experiment is performed}} - P[A] = \frac{X_1 + \ldots + X_n}{n} - \mu$. We have

$$\mathrm{E}[X] = \mathrm{E}\left[\frac{X_1 + \ldots + X_n}{n} - \mu\right] = \frac{n\,\mu}{n} - \mu = 0$$

and

$$\mathrm{Var}[X] = \mathrm{Var}\left[\frac{X_1 + \ldots + X_n}{n}\right] = \frac{n\,\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Therefore, $\mathrm{E}\left[X^2\right] = \mathrm{Var}[X] + \mathrm{E}[X]^2 = \frac{\sigma^2}{n}$. Using Chebyshev's Inequality with $k = 2$,

$$P[\,|X| \geq \varepsilon] \leq \frac{\sigma^2}{n\,\varepsilon^2} \xrightarrow{n\to\infty} 0\,.$$

■