Probability and Random Process

COSE382

PDF of a function of random variables

In general, PDF of a function of random variables is computed via CDF.

- **Function of a random variable:** Given a random variable X and a function $g : \mathbb{R} \to \mathbb{R}$. Let
$$Y := g(X)$$
T
$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \int_{\{x|g(x)\leq y\}} f_X(x)\mathrm{d}x$$
$$f_Y(y) = \frac{d}{dy}F_Y(y)$$

- **Function of two random variables:** Given random variables X and Y and a function $g : \mathbb{R}^2 \to \mathbb{R}$. Let
$$Z := g(X, Y)$$
$$F_Z(z) = \mathrm{P}(Z \leq z) = \mathrm{P}(g(X, Y) \leq z) = \iint_{\{x,y|g(x,y)\leq z\}} f_{XY}(x, y)\mathrm{d}x\mathrm{d}y$$
$$f_Z(z) = \frac{d}{dz}F_Z(z)$$

Example 8.1.3 (Chi-Square PDF). Let $X \sim \mathcal{N}(0,1)$ and

$$Y = X^2$$

Y is an example of a *Chi-Square* distribution (with degree of freedom 1). Find the PDF of Y

Solution: Starting form CDF again,

$$\begin{aligned} F_Y(y) &= P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\ &= 2\Phi(\sqrt{y}) - 1, \\ f_Y(y) &= \frac{d}{dy}F(y) = 2\varphi(\sqrt{y}) \cdot \frac{1}{2}y^{-1/2} = \varphi(\sqrt{y})y^{-1/2} \\ &= \frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2}, \quad y > 0 \end{aligned}$$

Example: Sum

Example (pdf of $Z = X + Y$) Let $Z = X + Y$

$$F_Z(z) = \mathrm{P}(X + Y \le z) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)dydx$$

$$f_Z(z) = \frac{d}{dz} \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)dydx = \int_{x=-\infty}^{\infty} \left[\frac{d}{dz} \int_{y=-\infty}^{z-x} f_{XY}(x,y)dy \right] dx$$

$$= \int_{x=-\infty}^{\infty} f_{XY}(x, z-x)dx$$

When X and Y are independent

$$f_Z(z) = \int_{x=-\infty}^{\infty} f(x) f_Y(z-x)dx,$$

becomes the convolution of f_X and f_Y.

Maximum of independent r.v.'s

Maximum Let $X = \max\{X_1, X_2, \cdots, X_n\}$, where $X_1, \cdots, X_n$ are independent.

$$\begin{aligned} F_X(x) &= P(X \le x) = P(\max\{X_1, X_2, \cdots, X_n\} \le x) \\ &= P(X_1 \le x \text{ and } X_2 \le x, \cdots \text{ and } X_n \le x) \\ &= \prod_{i=1}^{n} P(X_i \le x) = \prod_{i=1}^{n} F_{X_i}(x) \end{aligned}$$

If $\{X_i\}$ are i.i.d.,

$$\begin{aligned} F_X(x) &= F_{X_1}^n(x) \\ f_X(x) &= nF_{X_1}^{n-1}(x) f_{X_1}(x) \end{aligned}$$

Minimum of independent r.v.'s

Minimum Let $X = \min\{X_1, X_2, \cdots, X_n\}$, where $X_1, \cdots, X_n$ are independent.

$$
\begin{aligned}
F_X(x) = P(X \le x) = 1 - P(X > x) &= 1 - P\left(\min\{X_1, X_2, \cdots, X_n\} > x\right) \\
&= 1 - P\left(X_1 > x \text{ and } X_2 > x, \cdots \text{ and } X_n > x\right) \\
&= 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - \prod_{i=1}^{n}(1 - F_{X_i}(x))
\end{aligned}
$$

If $\{X_i\}$ are i.i.d.,

$$
\begin{aligned}
F_X(x) &= 1 - (1 - F_{X_1}(x))^n \\
f_X(x) &= n(1 - F_{X_1}(x))^{n-1} f_{X_1}(x)
\end{aligned}
$$

Example: Minimum of Exponentials

Find pdf of $X = \min(X_1, \cdots, X_n)$ where X_i are independent and $X_i \sim \text{Expo}(\lambda_i)$.

From $F_{X_i}(x) = 1 - e^{-\lambda_i x}$

$$F_X(x) = 1 - \prod_i (1 - (1 - e^{-\lambda_i x})) = 1 - \prod_i e^{-\lambda_i x} = 1 - e^{-\sum \lambda_i x}$$

$$f_X(x) = \left(\sum_i \lambda_i \right) e^{-(\sum_i \lambda_i) x}$$

Hence, $X \sim \text{Expo}(\lambda_1 + \cdots + \lambda_n)$

Change of variables

If g is invertible and differentiable, we have an easy way to compute PDF.

Theorem 8.1.1 (Change of variables in one dimension). Let X be a continuous r.v. with PDF f_X and let $Y = g(X)$, where g is differentiable and strictly increasing (or strictly decreasing). Then the PDF of Y is given by

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= f_X(x) \left| \frac{dx}{dy} \right|, \quad \text{where } x = g^{-1}(y) \end{aligned}$$

where dx/dy is the derivative of g^{-1} evaluated at y. Form the inverse function theorem,

$$\frac{dx}{dy} = \frac{d}{dy} g^{-1}(y) = \frac{1}{\frac{d}{dx} g(x)} = \frac{1}{\frac{dy}{dx}}$$

You can compute either dy/dx or dx/dy at your convenience (mostly dy/dx is easy).

Proof: For $Y = g(X)$, consider y with $y = g(x)$. To show is

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| .$$

When g is strictly increasing, $\frac{dx}{dy} > 0$ and

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(x), \\ f_Y(y) &= \frac{d}{dy} F_X(x) = f_X(x) \frac{dx}{dy}. \end{aligned}$$

When g is strictly decreasing, $\frac{dx}{dy} < 0$ and

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(g(X) \le y) = P(X \ge g^{-1}(y)) = 1 - F_X(g^{-1}(y)) = 1 - F_X(x), \\ f_Y(y) &= \frac{d}{dy}(1 - F_X(x)) = -f_X(x) \frac{dx}{dy}. \end{aligned}$$

Together, we have

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| .$$

Example 8.1.2 (Log-Normal PDF). Y is called Log-Normal r.v., if $Y = e^X$ for $X \sim \mathcal{N}(0,1)$. Find the PDF of Y.

Solution: $g(x) = e^x$ is strictly increasing. Let $y = e^x$, so $x = g^{-1}(y) = \ln y$

$$\frac{dx}{dy} = \frac{d}{dy}\ln y = \frac{1}{y} = \frac{1}{e^x}$$

On the other hand,

$$\frac{dy}{dx} = \frac{d}{dx}e^x = e^x = y$$

Note that dy/dx is easy to compute. Therefore,

$$\begin{aligned} f_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| = \varphi(x)\left|\frac{1}{e^x}\right| = \varphi(\ln y)\frac{1}{y}, \quad \text{for } y > 0 \\ &= \frac{1}{2\pi}y^{-1}\exp(-\frac{1}{2}(\ln y)^2) \quad \text{for } y > 0. \end{aligned}$$

Example 8.1.4 (PDF of a location-scale transformation). Let X have PDF f_X, and let

$$Y = a + bX,$$

with $b \neq 0$. Then, $g(x) = a + bx$ and $x = (y - a)/b$ and $\dfrac{dx}{dy} = 1/b$. The PDF of Y is

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = f_X\left(\frac{y - a}{b} \right) \frac{1}{|b|}.$$

Theorem 8.1.5 (Change of variables). Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a continuous random vector with joint PDF $f_{\mathbf{X}}$, and let $\mathbf{Y} = g(\mathbf{X})$ where g is an invertible function from $\mathbb{R}^n \to \mathbb{R}^n$.

Let $\mathbf{y} = g(\mathbf{x})$, and suppose that all the partial derivatives $\frac{\partial x_i}{\partial y_i}$ exist and are continuous, so we can form the *Jacobian matrix*:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}.$$

Also assume that the determinant of the Jacobian matrix is never 0. Then the joint PDF of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \left| \det \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right) \right| f_{\mathbf{X}}(g^{-1}(\mathbf{y})) = \frac{1}{\left| \det \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right) \right|} f_{\mathbf{X}}(g^{-1}(\mathbf{y})),$$

$\left| \det \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right) \right|$ is the absolute value of the determinant of the Jacobian matrix $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$.

Example 8.1.7 (Box-Muller). Let $U \sim \text{Unif}(0, 2\pi)$, and let $T \sim \text{Expo}(1)$ be independent of U. Define $X = \sqrt{2T}\cos U$ and $Y = \sqrt{2T}\sin U$. Find the joint PDF of (X, Y) and marginal distributions. Are they independent?

Solution: The joint PDF of (U, T) is $f_{U,T}(u,t) = \dfrac{1}{2\pi}e^{-t}$, for $u \in (0, 2\pi)$ and $t > 0$. The transform $(T, U) \to (X, Y)$ is invertible, sine it is a coordinate change from the polar to the rectangular:

$$X^2 + Y^2 = 2T(\cos^2 U + \sin^2 U) = 2T$$

The Jacobian matrix is given as

$$\frac{\partial(x,y)}{\partial(u,t)} = \begin{pmatrix} -\sqrt{2t}\sin u & \frac{1}{\sqrt{2t}}\cos u \\ \sqrt{2t}\cos u & \frac{1}{\sqrt{2t}}\sin u \end{pmatrix}$$

has absolute determinant $|-\sin^2 u - \cos^2 u| = 1$. Letting $x = \sqrt{2t}\cos u$, $y = \sqrt{2t}\sin u$, we have

$$\begin{aligned} f_{X,Y}(x,y) &= f_{U,T}(u,t) \cdot \left| \det \frac{\partial(u,t)}{\partial(x,y)} \right| = \frac{1}{2\pi}e^{-t} \cdot 1 = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)} \\ &= \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}}e^{-y^2/2} \end{aligned}$$

X and Y are i.i.d. $\mathcal{N}(0, 1)$. This result is called the Box-Muller method for generating Normal r.v.s.

(PDF of jointly Gaussian). Now we are ready to prove joint PDF of jointly Gaussian. Let $X = [X_1, \cdots, X_n]^T$ be a jointly Gaussian with $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_X)$. Then the joint PDF of X is given by

$$f_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\det \Sigma_X|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma_X^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Proof: Since Σ is a covariance matrix, there is a symmetric square root matrix of Σ, $\Sigma^{1/2}$, such that $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. Let $\mathbf{z} = \Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$, where $\Sigma^{-1/2}$ is the inverse of $\Sigma^{1/2}$ ($\Sigma^{1/2}$ and $\Sigma^{-1/2}$ always exist!). Then the covariance of $\mathbf{z}$ is

$$\Sigma_Z = E(\mathbf{z}\mathbf{z}^T) = \Sigma^{-1/2}E(\mathbf{x}\mathbf{x}^T)\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma\Sigma^{-1/2} = I$$

Hence, $\mathbf{z}$ is i.i.d. Gaussian vectors of $\mathcal{N}(0,1)$ and

$$f_Z(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right)$$

From $\mathbf{z}$ we get $\mathbf{x}$ back from $\mathbf{x} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}$. Sine the Jacobian of $d\mathbf{x}/d\mathbf{z} = \Sigma_X^{1/2}$, and $\det(\Sigma_X^{1/2}) = \det(\Sigma_X)^{1/2}$,

$$f_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\det \Sigma_X|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma_X^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Gamma function

Definition 8.4.1 (Gamma function). The *gamma function* Γ is defined by

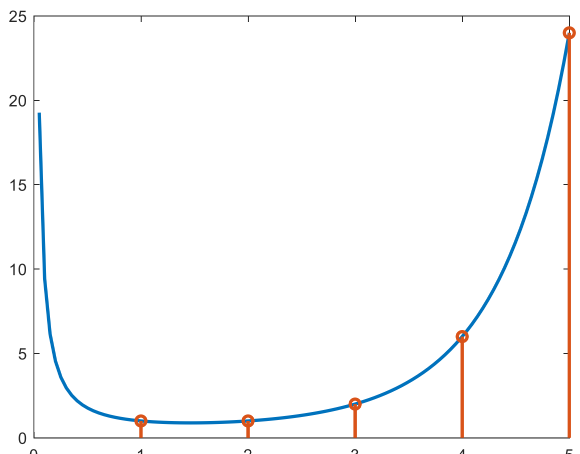
$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx = \int_0^\infty \lambda^a x^{a-1} e^{-\lambda x} dx \text{ (for any } \lambda > 0),$$

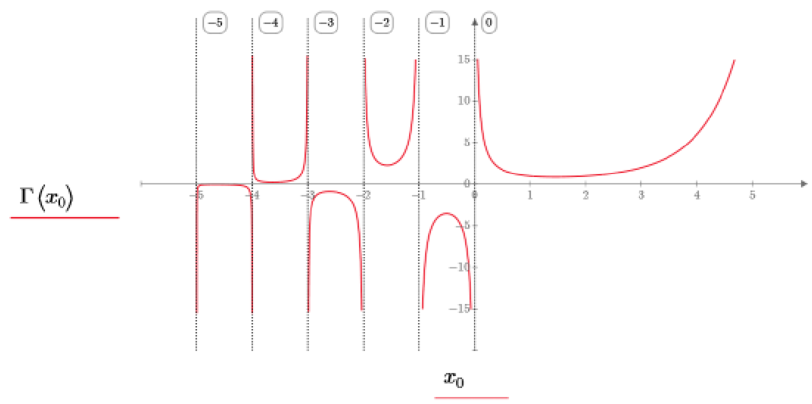
for real numbers $a > 0$.

- $\Gamma(a+1) = a\Gamma(a)$ for all $a > 0$. This follows from integration by parts:

$$\Gamma(a+1) = \int_0^\infty x^a e^{-x} dx = -x^a e^{-x}\Big|_0^\infty + a\int_0^\infty x^{a-1} e^{-x} dx = 0 + a\Gamma(a).$$

- $\Gamma(n) = (n-1)!$ if n is a positive integer.





Definition 8.3.1 (Beta distribution). $X \sim \text{Beta}(a, b)$: A r.v. X is said to have the *Beta distribution* with parameters a and b, $a, b > 0$, if its PDF is

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1,$$

where the constant $\beta(a, b)$ is chosen to make the PDF integrate to 1;

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

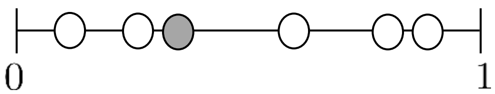
An integral of this form is called a beta integral. When a, b are integers,

$$\beta(a,b) = \frac{1}{(a+b-1)\binom{a+b-2}{a-1}} = \frac{(a-1)!(b-1)!}{(a+b-1)!} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

from the fact (Bayes' billiards)

$$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dp = \frac{1}{n+1} \implies \int_0^1 x^k (1-x)^{n-k} dp = \frac{1}{(n+1)\binom{n}{k}}$$

Bayes' Billiards



Start with $n+1$ balls, n white and 1 gray. Randomly throw each ball on to the unit interval $(0,1)$. Let X be the number of white balls to the left of gray ball. Find $P(X=k)$.

1. We use LOTP by conditioning on the position of the gray ball. Let G be the position of the gray ball. Conditional on $G=p\in[0,1]$, the random variable X has $\text{Bin}(n,p)$ distribution. Since $G\sim$ Unif(0,1),

$$P(X=k)=\int_0^1 P(X=k\mid G=p)f_G(p)\mathrm{d}p=\int_0^1\binom{n}{k}p^k(1-p)^{n-k}\ \mathrm{d}p$$

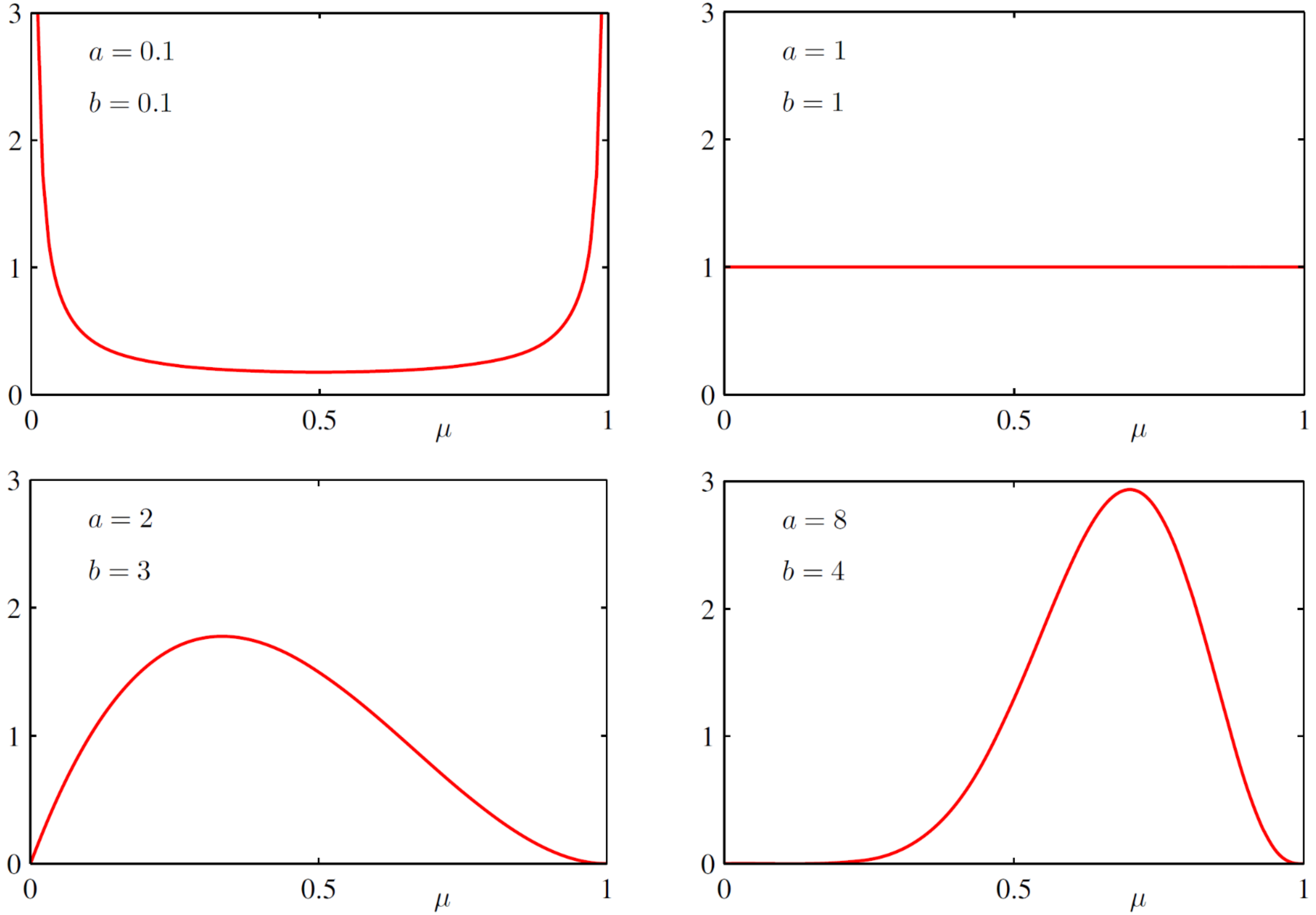
2. On the other hand: Start with $n+1$ balls, all white. Randomly throw each ball onto to the unit interval. Then choose one ball at random and paint it gray. The event $\{X=k\}$ is equivalent to the event "the $(k+1)$ th ball from the left is gray". Therefore

$$P(X=k)=\frac{1}{n+1}$$

Since X has the same distribution in both cases we have

$$\int_0^1\binom{n}{k}x^k(1-x)^{n-k}\ \mathrm{d}x=\frac{1}{n+1},\quad k\in\{0,1,\cdots,n\}.$$

Beta distribution



Example

- Example:

Statisticians say that the probability of success that someone would agree to go on a date with you follows a Beta($a = 2$, $b = 8$). What is the probability that your chance of having a date will be greater than 50% ?

Solution:

$$P(X > 0.5) = \frac{1}{\beta(2,8)} \int_{x=1/2}^{1} x(1-x)^7 dx = \frac{1!7!}{9!} \int_{x=1/2}^{1} x(1-x)^7 dx \approx 0.01953$$

Statisticians say that the probability of H of a coin manufactured in a factory follows a Beta($a = 2$, $b = 8$). What is the probability that your chance of having H by tossing the coin will be greater than 50% ?

Solution:

$$P(X > 0.5) = \frac{1}{\beta(2,8)} \int_{x=1/2}^{1} x(1-x)^7 dx = \frac{1!7!}{9!} \int_{x=1/2}^{1} x(1-x)^7 dx \approx 0.01953$$

Mean and Variance of Beta distr.

- We will shortly prove that

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

holds not just for integers but for all real $a, b > 0$.

- Then for $X \sim \text{Beta}(a,b)$,
  - The $n-th$ moment is

$$E[X^n] = \frac{1}{\beta(a,b)} \int_0^1 x^n x^{a-1}(1-x)^{b-1}\ dx = \frac{1}{\beta(a,b)} \int_0^1 x^{n+a-1}(1-x)^{b-1}\ dx$$
$$= \frac{\beta(a+n,b)}{\beta(a,b)} = \frac{\Gamma(a+n)\Gamma(b)}{\Gamma(a+b+n)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

  - The mean and variance are

$$E[X] = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{a}{a+b}$$
$$Var(X) = E(X^2) - E(X)^2 = \frac{\beta(a+2,b)}{\beta(a,b)} - \left(\frac{a}{a+b}\right)^2$$
$$= \frac{ab}{(a+b)^2(a+b+1)}$$

Gamma distribution

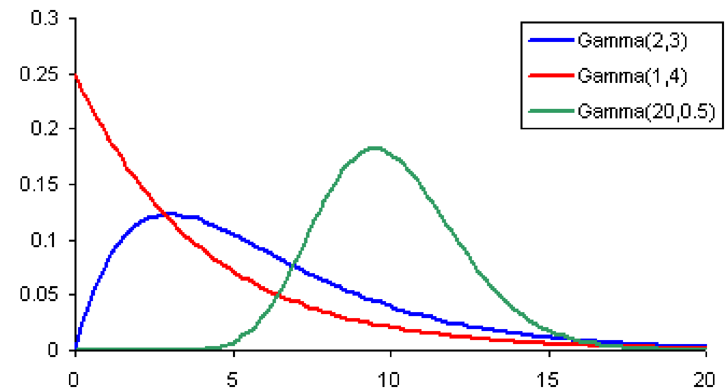
Definition 8.4.2 (Gamma distribution). An r.v. X has the *Gamma distribution* with parameters a and 1, denoted $X \sim \text{Gamma}(a, 1)$, if its PDF is

$$f_X(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x}, \quad x > 0.$$

$Y = \frac{1}{\lambda} X$ is called the $\text{Gamma}(a, \lambda)$ distribution:

$$\begin{aligned} f_Y(y) &= f_X(x) \left| \frac{dx}{dy} \right| = \frac{1}{\Gamma(a)} (\lambda y)^{a-1} e^{-\lambda y} \lambda \\ &= \frac{1}{\Gamma(a)} \lambda (\lambda y)^{a-1} e^{-\lambda y} \end{aligned}$$

- $\text{Gamma}(1, \lambda) = \lambda e^{-\lambda y} = \text{Expo}(\lambda)$



Theorem 8.4.3 Let $X_1, \dots, X_n$ be i.i.d. Expo(λ). Then

$$X_1 + \cdots + X_n \sim \text{Gamma}(n, \lambda).$$

Proof. The MGF of an Expo(λ) r.v. is $\frac{\lambda}{\lambda - t}$ for $t < \lambda$, so the MGF of $X_1 + \cdots + X_n$ is

$$M_n(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

for $t < \lambda$. Let $Y \sim \text{Gamma}(n, \lambda)$;

$$\begin{aligned}
E(e^{tY}) &= \int_0^\infty e^{ty} \frac{1}{\Gamma(n)} \lambda(\lambda y)^{n-1} e^{-\lambda y} dy \\
&= \frac{\lambda^n}{(\lambda - t)^n} \int_0^\infty \frac{1}{\Gamma(n)} (\lambda - t)((\lambda - t)y)^{n-1} e^{-(\lambda - t)y} dy \\
&= \left(\frac{\lambda}{\lambda - t}\right)^n
\end{aligned}$$

Hence, $X_1 + \cdots + X_n \sim \text{Gamma}(n, \lambda)$.

Example

- Example:

  A device is consisted of independently working three components and the life time of each component has exponential distribution with the averaged life time 1 year. The device is out of order when all of the three components are. What is the probability that the device is working more than 3 years?

  Solution: Let X denote the life time of the device. Then $X \sim \text{Gamma}(3, 1)$.

$$P(X > 3) = \frac{1}{\Gamma(3)} \int_{x=3}^{\infty} x^2 e^{-x} dx \approx 0.3232$$

Mean and Variance of Gamma distribution

The nth moment of Gamma(a,1).

$$E(X^n) = \int_0^\infty \frac{x^n}{\Gamma(a)} x^{a-1} e^{-x} dx = \frac{1}{\Gamma(a)} \int_0^\infty x^{a+n-1} e^{-x} dx = \frac{\Gamma(a+n)}{\Gamma(a)} = (a+n-1)\cdots a$$

Mean of Gamma$(a,1)$: $E(X) = a$.

Variance of Gamma$(a,1)$: $\mathrm{Var}(X) = E(X^2) - EX^2 = (a+1)a - a^2 = a$.

For $Y = X/\lambda \sim$ Gamma(a,λ);

$$E(Y) = \frac{1}{\lambda} E(X) = \frac{a}{\lambda},$$
$$\mathrm{Var}(Y) = \frac{1}{\lambda^2} \mathrm{Var}(X) = \frac{a}{\lambda^2},$$
$$E(Y^n) = \frac{1}{\lambda^c} E(X^n) = \frac{1}{\lambda^n} \cdot \frac{\Gamma(a+n)}{\Gamma(a)}.$$

Beta-Gamma Connection

Given two independent Gamma distributions with the same λ,

$$X \sim \text{Gamma}(a, \lambda), \quad Y \sim \text{Gamma}(b, \lambda),$$

Let $T = X + Y$ and $W = X/T$. What is the joint PDF of (T, W) ?

Solution: $X = TW$ and $Y = T(1 - W)$. Hence the Jacobian matrix is

$$\frac{\partial(x,y)}{\partial(t,w)} = \begin{pmatrix} w & t \\ 1-w & -t \end{pmatrix}, \quad \text{and} \quad \left|\frac{\partial(x,y)}{\partial(t,w)}\right| = t$$

$$\begin{aligned}
f_{T,W}(t,w) &= f_{X,Y}(x,y)\left|\frac{\partial(x,y)}{\partial(t,w)}\right| = \frac{1}{\Gamma(a)}\lambda(\lambda x)^{a-1}e^{-\lambda x} \cdot \frac{1}{\Gamma(b)}\lambda(\lambda y)^{b-1}e^{-\lambda y} \cdot t \\
&= \frac{1}{\Gamma(a)}\lambda(\lambda tw)^{a-1}e^{-\lambda tw} \cdot \frac{1}{\Gamma(b)}\lambda(\lambda t(1-w))^{b-1}e^{-\lambda t(1-w)} \cdot t \\
&= \frac{1}{\Gamma(a)\Gamma(b)}w^{a-1}(1-w)^{b-1}\lambda(\lambda t)(\lambda t)^{a-1+b-1}e^{-\lambda tw-\lambda t(1-w)} \\
&= \underbrace{\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}w^{a-1}(1-w)^{b-1}\right)}_{f_W(w)}\underbrace{\left(\frac{1}{\Gamma(a+b)}\lambda(\lambda t)^{a+b-1}e^{-\lambda t}\right)}_{f_T(t)}
\end{aligned}$$

Beta-Gamma Connection

- **Beta-Gamma connection:** For independent $X \sim \text{Gamma}(a, \lambda), Y \sim \text{Gamma}(b, \lambda)$

$$\begin{aligned} T &= X + Y && \sim \text{Gamma}(a+b, \lambda) \\ W &= \frac{X}{T} = \frac{X}{X+Y} && \sim \text{Beta}(a, b) \end{aligned}$$

  T and W are independent

- **Gamma+Gamma:**

$$\text{Gamma}(a, \lambda) + \text{Gamma}(b, \lambda) = \text{Gamma}(a+b, \lambda), \text{ for any } a, b > 0$$

- **Beta integral:**

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}dx, \text{ for any } a, b > 0$$

Chi-squared with n degree of freedom

Definition: (Chi-squared with n degree of freedom). The chi-squared distribution with n degrees of freedom (or chi- squared (n), or $\chi^2(n)$) is the distribution of the random variable

$$z = X_1^2 + X_2^2 + \cdots + X_n^2,$$

where $X_i \sim \mathcal{N}(0,1)$

Theorem. (Pdf of $\chi^2(n)$) Chi-squared is a Gamma distribution

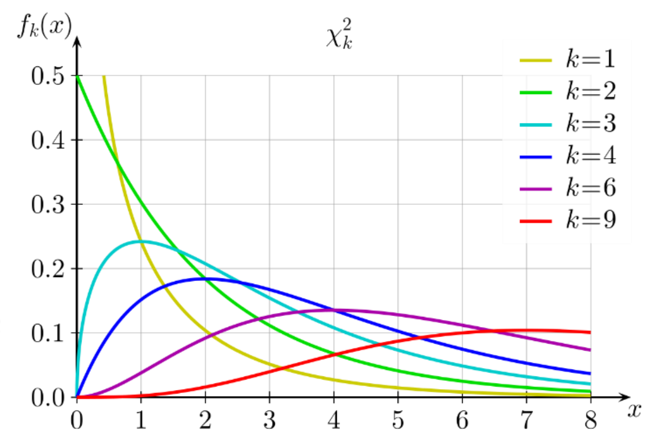
Proof: From thg ealier example we have shown that for $Z \sim \chi^2(1)$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} z^{-1/2} e^{-z/2}, \quad z > 0$$

This implies that $\chi^2(1) \sim \mathrm{Gamma}(1/2, 1/2)$. Hence,

$$\chi^2(n) \sim \mathrm{Gamma}(n/2, 1/2)$$

$$f_Z(z) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{(n/2)-1} e^{-z/2}$$



Gaussian Sample Variance

Let $X_1, \ldots, X_n$ be i.i.d. $N\left(\mu, \sigma^2\right)$. Define

$$\bar{X} = \frac{1}{n}\left(X_1 + \cdots + X_n\right) \text{ and } S = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$\bar{X}$ is an estimate of μ, called sample mean and $\hat{\sigma}^2 := \frac{1}{n-1}S$ is an estimate of σ^2, called the sample variance. We can show $S \sim \sigma^2\chi^2(n-1)$ so that $E(\hat{\sigma}^2) = \sigma^2$ and furthermore $\hat{\sigma}^2$ and $\bar{X}$ are independent.

Proof: We rewrite $\frac{1}{\sigma^2}S = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2$ as

$$\frac{1}{\sigma^2}S = \left(\frac{X_1 - X_2}{\sigma\sqrt{2}}\right)^2 + \left(\frac{X_1 + X_2 - 2X_3}{\sigma\sqrt{2\cdot 3}}\right)^2 + \left(\frac{X_1 + X_2 + X_3 - 3X_4}{\sigma\sqrt{3\cdot 4}}\right)^2 + \cdots + \left(\frac{X_1 + X_2 + \cdots + X_{n-1} - (n-1)X_n}{\sigma\sqrt{(n-1)n}}\right)^2$$

Each of the $n-1$ expressions within brackets has the standard normal distribution. Furthermore, the expressions within brackets are all independent of one another and are also all independent of $\bar{X}$ It follows that S is independent of $\bar{X}$. Therefore, by the definition of the chi-squared distribution, $S/\sigma^2 \sim \chi^2(n-1)$

t-distribution

The t distribution with n degrees of freedom (or Student (n), or $t(n)$), is the distribution of the random variable

$$T = \frac{X}{\sqrt{\left(X_1^2 + X_2^2 + \cdots + X_n^2\right)/n}}$$

where $X, X_1, \ldots, X_n$ are i.i.d., each with the standard normal distribution $\mathcal{N}(0,1)$ Equivalently,

$$T = \frac{X}{\sqrt{Y/n}},$$

where $Y \sim \chi^2(n)$ independent from X. The pdf of T is given as

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

Sketch of Proof: X and Y have joint pdf given by

$$f_{X,Y}(x,y) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \cdot \frac{y^{(n/2)-1}e^{-y/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \quad \text{for } y > 0$$

Consider a transform $(X,Y) \to (T,V)$, where $T = X/\sqrt{Y/n}$ and $V = Y$. Then $X = T\sqrt{V/n}$. We compute $f_{T,V}(t,v)$ from $f_{X,Y}(x,y)$ using change of variables. The determinant of Jacobian is given by

$$\det \frac{\partial(t,v)}{\partial(x,y)} = \det \begin{pmatrix} \frac{1}{\sqrt{y/n}} & 0 \\ \frac{-x\sqrt{n}}{y^{3/2}} & 1 \end{pmatrix} = \frac{\sqrt{n}}{\sqrt{y}}$$

Hence,

$$f_{T,V}(t,v) = \frac{1}{\sqrt{\pi}\Gamma(n/2)} \frac{1}{2^{(n+1)/2}} \frac{1}{\sqrt{n}} v^{(n+1)/2-1} e^{-(v/2)\left(1+t^2/n\right)}$$

for $v > 0$. Note that in terms of v, $f_{T,V}$ contains a Gamma function with $a = (n+1)/2 - 1$ and $\lambda = \frac{1}{2(1+t^2/n)}$.

Finally, we compute the marginal density of T :

$$f_T(t) = \int_{-\infty}^{\infty} f_{T,V}(t,v)dv = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \frac{1}{\sqrt{n}}$$

Mean estimation and t-distribution

- Consider an i.i.d. distribution $X_1, \cdots, X_n$ with $X_i \sim \mathcal{N}(\mu, \sigma^2)$. The normalized sample mean,

$$Z = \frac{(\bar{X} - \mu)/\sqrt{n}}{\sigma} \sim \mathcal{N}(0,1).$$

However, when the variance σ is unknown and approximated by the sample variance $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, the approximated normalized mean

$$T = \frac{(\bar{X} - \mu)/\sqrt{n}}{\sqrt{\hat{\sigma}^2}} \sim t(n-1),$$

since $\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \sigma^2\chi(n-1)$.

- Using $t(n)$ instead $\mathcal{N}(0,1)$ became popular by a paper written by a pseudonym "Student," a statistician working in Guinness Brewery, William Sealy Gosset.