The probability that k flips are needed until a heads appears with a coin that has probability p of coming up heads is

$$P(X=k) = (1-p)^{k-1}p \quad \text{for} \quad k \geq 1$$

Def: X is called a geometric random variable with parameter p. We write $X \sim \text{Geo}(p)$.

Def: A valid probability function satisfies the conditions that all probabilities are nonnegative and sum to 1.

Note: $P(X \geq k) = (1-p)^{k-1}$

Def: The expected value of X is a weighted average of all possible values.

$$\mathbb{E}(X) := \sum_k kP(X=k)$$

Note: For $X \sim \text{Geo}(p)$, $\mathbb{E}(x) = \frac{1}{p}$.

Ⓠ: A coin with heads probability p is flipped n times. Let X denote the number of heads obtained. Find $P(X=k)$.

Ⓐ: $\binom{n}{k}p^k(1-p)^{n-k}$, $k = 0, \ldots, n$

Def: X is called a binomial random variable with parameters n and p. We write $X \sim \text{Bin}(n,p)$. We verify this is a probability function with the binomial theorem.

Note: For $X \sim \text{Bin}(n,p)$, $\mathbb{E}(x) = np$

Suppose n is large, p is small, and $k \ll n$. Let $\lambda = np$. Then

$$\begin{aligned}
P(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\ldots(n-k+1)}{k!} p^k (1-p)^{n-k} \\
&\approx \frac{n^k}{k!} p^k (1-p)^{n-k} && (k \ll n) \\
&\approx \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n && (n-k \approx n) \\
&\approx \frac{\lambda^k}{k!} e^{-\lambda} && (n \to \infty)
\end{aligned}$$

Def: A Poisson random variable $X \sim \text{Poi}(\lambda)$ has

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots$$

and typically counts rare events. (n large, p small, $X \sim \text{Bin}(n,p)$) Here, λ denotes the average number.

Note: This is a probability function since

$$\begin{aligned}\sum_{k \geq 0} P(X=k) &= \sum_{k \geq 0} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} e^{\lambda} \\ &= 1\end{aligned}$$

Note: $\mathbb{E}(X) = \lambda$ since

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k \geq 0} k P(X=k) \\ &= \sum_{k \geq 1} k \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \lambda e^{-\lambda} \sum_{k \geq 1} \frac{\lambda^{k-1}}{k-1!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda\end{aligned}$$

Def: For events A and B, the probability that A occurs given that B occurs is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Note: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Def: When A and B are disjoint or mutually exclusive, $P(A \cap B) = 0$, $P(A \cup B) = P(A) + P(B)$, and $P(A|B) = 0$.

Def: When A and B are independent, $P(A \cap B) = P(A)P(B)$.

Note: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

Thm: (The Law of Total Probability) $P(A) = \sum_{i=1}^{N} P(A|B_i)P(B_i)$

Note: $\mathbb{E}[X]$ sometimes doesn't exist (i.e. undefined or infinite)

Ex: $X = \begin{cases} 2 & \text{w/ pr } 1/2 \\ 4 & 1/4 \\ 8 & 1/8 \\ \vdots & \end{cases}$

$E[X]$ is infinite

$X = \begin{cases} 2 & \text{w/ pr } 1/4 \\ -2 & 1/4 \\ 4 & 1/8 \\ -4 & 1/8 \\ \vdots & \end{cases}$

$E[X]$ is undefined (sum doesn't converge)

Thm: $E[aX+b] = aE[X]+b$

Pf: $E[aX+b] = \sum (ax+b)P(X=x) = a\sum xP(X=x) + b\sum P(X=x) = aE[X]+b$

Def: Suppose X is a random variable with mean μ. The variance of X is the average squared distance X is from μ. It measures a random variable's volatility.

$$Var[X] = E[(X-\mu)^2] = E[(X-E[X])^2]$$

Thm: $Var(aX+b) = a^2 Var(X)$

Pf: Let $Y = aX+b$. $Var(Y) = E[(Y-\mu_Y)^2] = E[(aX+b-a\mu_Y-b)^2] = E[a^2(X-\mu_X)^2] = a^2 Var(X)$

Thm: $Var(X) = E[X^2] - E[X]^2$

Pf: $Var(X) = E[(X-\mu)^2] = \sum (x-\mu)^2 P(X=x) = \sum x^2 P(X=x) - \sum 2x\mu P(X=x) + \sum \mu^2 P(X=x)$

$= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2$

Def: The standard deviation of X is $SD(X) = \sqrt{Var(X)}$ and is often denoted σ.

Note: $X \sim Bin(n,p) \Rightarrow Var(X) = np(1-p)$

$X \sim Poi(\lambda) \Rightarrow Var(X) = \lambda$

Pf: $Var(X) = E[X^2] - E[X]^2 = E[X^2] - \lambda^2$

$$E[X^2] = E[X(X-1)+X] = E[X(X-1)] + E[X] = \sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x e^{-\lambda}}{x!} + \lambda \overset{\substack{\text{terms are 0} \\ \text{when } x=0,1 \\ \downarrow}}{=} \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}e^{-\lambda}}{(x-2)!} + \lambda = \lambda^2 + \lambda$$

$Var(X) = E[X^2] - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Def: A probability density function f_X describes a continuous random variable. For any $a \le b$, $P(a \le X \le b) = \int_a^b f_X(x)dx$ and satisfy $f_X(x) \ge 0$ for all x and $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Note: $P(X=a) = \int_a^a f_X(x)dx = 0$

Def: For continuous random variables, $E[X] = \int_x x f_X(x)dx$

Note: $E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$

Def: The exponential distribution is $X \sim Expo(\lambda)$ $(\lambda > 0)$ with PDF

$$f_X(X) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

We verify this is a PDF.

(i) $f_X(x) = \lambda e^{-\lambda x} \geq 0$, $\forall x$

(ii) $\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0 \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^{\infty} = 1$

☾

Note: For any $t > 0$, $P(X > t) = \int_t^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_t^{\infty} = e^{-\lambda t}$

Note: $E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{1}{\lambda}$ $\quad$ $Var(X) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Note: In practice, X models times between rare events. This is intimately related to the Poisson random variable, using the same λ.

Ex: If $X \sim Poi(\lambda)$, $E(X) = \lambda$, say $E(X) = 5$ cars/hour, if Y is the time between cars, $Y \sim Expo(\lambda)$ with $E(Y) = \frac{1}{\lambda} = 12$ minutes.

Def: The uniform distribution is $X \sim U(a,b)$ with PDF $f_X(x) = \frac{1}{b-a}$, $a < x < b$. It's easy to verify this is a PDF, $E(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.

Note: $P(c < X < d) = \frac{d-c}{b-a}$ for $a < c < d < b$

Def: The Cauchy distribution is $X \sim Cau(0,1)$ with PDF $f_X(x) = \frac{1}{\pi(1+x^2)}$. Note

☾ $f_X(x) \geq 0$, $\forall x$ and $\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)}\,dx = \frac{1}{\pi} \tan^{-1}(x)\Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = 1$.

Note: $E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}\,dx = \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)}\,dx + \int_0^{\infty} \frac{x}{\pi(1+x^2)}\,dx = \frac{1}{2\pi}\int_{\infty}^{1} \frac{1}{u}\,du + \frac{1}{2\pi}\int_1^{\infty} \frac{1}{u}\,du$

$= \frac{1}{2\pi} \ln u \Big|_{\infty}^{1} + \frac{1}{2\pi} \ln u \Big|_1^{\infty} = -\infty + \infty$ which is undefined.

Def: The normal distribution is $X \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2, and $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ We can verify that $f_X(x)$ is a PDF, $E(X) = \mu$, and $Var(X) = \sigma^2$.

Def: The standard normal distribution is $X \sim N(0,1)$. Traditionally, its PDF is denoted by $\phi(x)$. We have $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Note: About 95% of the normal distribution is between $[\mu - 1.96\sigma, \mu + 1.96\sigma]$.

Note: $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx \overset{\substack{y = \frac{x-\mu}{\sigma} \\ dy = dx/\sigma}}{=} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\,dy$ so it remains to show $I = \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\,dy = \sqrt{2\pi}$.

☾ We see $I^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\,dy\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy = \int_0^{2\pi}\int_0^{\infty} r e^{-\frac{1}{2}r^2}\,dr\,d\theta$

$= \int_0^{2\pi} -e^{-\frac{1}{2}r^2}\Big|_0^{\infty}\,d\theta = \int_0^{2\pi} d\theta = 2\pi$

so $I = \sqrt{2\pi}$ as desired, and $N(\mu, \sigma^2)$ is a PDF.

Def: For any random variable X, we define the cumulative distribution function to be $F_X(x) = P(X \le x)$. If X is discrete, then $F_X(x) = \sum_{k \le x} P(X=k)$. If X is continuous with PDF f_X, then $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$.

Ex: Suppose $X = \begin{cases} -1 & \text{w/ prob } 1/2 \\ 0 & 1/3 \\ 1 & 1/6 \end{cases}$ and $Y = X^2+5$. The PMF of Y is $Y = \begin{cases} 5 & \text{w/ prob } 1/3 \\ 6 & 2/3 \end{cases}$

Ex: Suppose X is the Centigrade temperature at some point and has PDF $f_X(x) = cx^2$, $0 < x < 30$. Find the PDF of $Y = \frac{9}{5}X + 32$. Since $0 < X < 32$, $32 < Y < 86$. Also, $F_Y(y) = P(Y \le y) = P(\frac{9}{5}X + 32 \le y) = P(X \le \frac{5}{9}(y-32)) = F_X(\frac{5}{9}(y-32))$. Differentiating both sides gives $f_Y(y) = \frac{5}{9} f_X(\frac{5}{9}(y-32)) = c(\frac{5}{9})^3(y-32)^2$, $32 < Y < 86$.

Note: Suppose X has PDF $f_X(x)$, $a < x < b$, and $Y = g(X)$ is strictly increasing then g^{-1} exists and is also increasing. Note $a < x < b \Rightarrow g(a) < y < g(b)$. Then we have $F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$. Differentiating both sides gives $f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}(g^{-1}(y))$, $g(a) < y < g(b)$. Similarly if $Y = g(X)$ is strictly decreasing then $f_Y(y) = -f_X(g^{-1}(y)) \frac{d}{dy}(g^{-1}(y))$, $g(b) < y < g(a)$.

Ex: Suppose X has PDF f_X. Find the PDF of $Y = F_X$. Note $Y = F_X(x) = P(X \le x)$ so $0 < Y < 1$. Also, $F_Y(y) = P(Y \le y) = P(F_X(x) \le y) = P(X \le F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$. Thus, $f_Y(y) = 1$, $0 < y < 1$ and $Y \sim U(0,1)$.

Thm: Conversely, if $U \sim U(0,1)$ then $Y = F_X^{-1}(U)$ has the same PDF as X.

Pf: $F_Y(y) = P(Y \le y) = P(F_X^{-1}(U) \le y) = P(U \le F_X(y)) = F_X(y) \implies f_Y(y) = f_X(y)$

Note: Let $X \sim N(\mu, \sigma^2)$. Then $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{1}{2\sigma^2}(x-\mu)^2}$. Let $z = \frac{X-\mu}{\sigma}$. We want to find $f_Z(z)$. The feasible region is $-\infty < x < \infty \Rightarrow -\infty < \frac{x-\mu}{\sigma} = z < \infty$. Then, $F_Z(z) = P(Z \le z) = P(\frac{X-\mu}{\sigma} \le z) = P(X \le \sigma z + \mu) = F_X(\sigma z + \mu)$. Thus, $f_Z(z) = f_X(\sigma z + \mu) \cdot \sigma = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(\sigma z + \mu - \mu)^2} \sigma = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \phi(z)$. Therefore $Z \sim N(0,1)$. Any normal problem can be standardized to a standard normal problem.

Note: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ has no simple anti-derivative.

Prop: If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ then $Y \sim N(a\mu + b, a^2\sigma^2)$.

Prop: Any linear combination of independent normal random variables is also a normal random variable. Specifically, if $X_1, \ldots, X_n$ are independent normal random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ then $X_1 + X_2 + \ldots + X_n \sim N(\Sigma \mu_i, \Sigma \sigma_i^2)$ and $\Sigma a_i X_i + b \sim N(\Sigma a_i \mu_i + b, \Sigma a_i \sigma_i^2)$.

Note: Suppose X_1, X_2 are independent $N(\mu, \sigma^2)$. Consider $X_1 + X_2$ and $2X_1$. Are they the same? They are not since $X_1 + X_2 \sim N(2\mu, 2\sigma^2)$ and $2X_1 \sim N(2\mu, 4\sigma^2)$.

Def: Jointly distributed random variables can be defined by a joint probability mass function in the discrete case.

Ex:

x \ Y	1	2	3
0	.1	.2	.3
1	.3	.1	0

$P(X=0 \text{ and } Y=2) = P((X,Y)=(0,2)) = 0.2$

Def: The marginal probability mass functions are the probabilities when considering only one random variable.

Note: We can't reconstruct a probability mass function given marginal probability mass functions but if the random variables are independent then we can.

Def: In general, a joint probability mass function must satisfy

(i) $P(X=x, Y=y) \geq 0$, $\forall x,y$

(ii) $\sum_x \sum_y P(X=x, Y=y) = 1$

The marginal probability mass function is obtained by summing over the possible values of one random variable, i.e. $P(X=k) = \sum_y P(X=k, Y=y)$.

Def: For the continuous case, given two random variables X, Y, their joint probability density function $f_{X,Y}$ must satisfy

(i) $f_{X,Y}(x,y) \geq 0 \quad \forall x,y$

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$

Here, $P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy$

Def: X and Y are independent if $\forall x,y$ $P(X=x, Y=y) = P(X=x)P(Y=y)$.

Note: Suppose we were given $f_{x,y}$ but we want f_x. Then

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

Def: $P(A\mid B) = \frac{P(A\cap B)}{P(B)}$

Remark: The big picture is that the conditional distribution of Y given X is the distribution of Y and X over the distribution of X.

Ex:

X\Y	1	2	3	
0	.1	.2	.3	.6
1	.3	.1	.0	.4
	.4	.3	.3	

$P(Y=1)=.4$ $\quad$ But we have $\quad$ $P(Y=1\mid X=0)=\frac{1}{6}$

$P(Y=2)=.3$ $\quad$ the following $\quad$ $P(Y=2\mid X=0)=\frac{2}{6}$

$P(Y=3)=.3$ $\quad$ conditional distribution $\quad$ $P(Y=3\mid X=0)=\frac{3}{6}$

Note: In general, the PMF for Y given $X=x$ is $P(Y=y\mid X=x) = \frac{P(Y=y \text{ and } X=x)}{P(X=x)}$ ← joint PMF (fixed x) ← marginal of X at x

Ex: Continuous case: $f_{X,Y}(x,y)=\frac{5}{2}x^2y$, $0\le y\le 2x\le 2$. Then $f_{Y|X}(y\mid\frac{1}{2}) = \frac{5}{2}(\frac{1}{2})^2y/[\int_0^1 \frac{5}{8}y\,dy]$ $= 2y$, $0\le y\le 1$. We can also compute $P(\frac{1}{2}\le Y\le\frac{3}{2}\mid X=\frac{1}{2}) = \int_{1/2}^1 2y\,dy = \frac{3}{4}$.

Note: In general, $f_{Y|X}(y\mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$.

Note: When X and Y are independent, $f_{Y|X}(y\mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y)$.

Ex: Suppose $X\sim U(0,1)$. After X is chosen, Y is chosen with $Y\sim U(x,1)$. Find the PDF of Y. Note $f_X(x)=1$, $0<x<1$, and $f_{Y|X=x}=\frac{1}{1-x}$, $x<y<1$. Therefore, $f_{X,Y}(x,y)=\frac{1}{1-x}$, $0<x<y<1$. Then $f_Y(y)=\int_0^y\frac{dx}{1-x} = -\ln(1-x)\Big|_0^y = \ln\left(\frac{1}{1-y}\right)$, $0<y<1$.

Def: A multivariable distribution is a distribution with more than 2 variables.

Ex: The multinomial distribution is the multivariate case of the binomial distribution.

Note: If $X_1,\dots,X_N$ have joint PDF $f_{X_1,\dots,X_N}$ then $P((X_1,\dots,X_n)\in A) = \int\cdots\int_A f_{X_1,\dots,X_N}(x_1,\dots,x_N)\,dx_1\dots dx_N$. The marginal density of X_1 is obtained by fixing X_1 and integrating out $X_2,\dots,X_N$.

Note: $X_1,\dots,X_N$ are independent if $f_{X_1,\dots,X_N}(x_1,\dots,x_N)=\prod f_{X_i}(x_i)$, $\forall x_i$.

Ex: Suppose $X_1,\dots,X_N$ are independent exponential random variables each with parameter λ, i.e. $X_1,\dots,X_N$ are i.i.d. (independent and identically distributed). Thus, X_i has PDF $f_{X_i}(x_i)=\lambda e^{-\lambda x_i}$, $x_i\ge 0$. Since $X_1,\dots,X_N$ are independent, their joint PDF is $f_{X_1,\dots,X_N}(x_1,\dots,x_N)=\prod f_{X_i}(x_i) = \prod\lambda e^{-\lambda x_i} = \lambda^N e^{-\lambda(x_1+\dots+x_N)}$, $x_1,\dots,x_N>0$.

Note: Given the joint PDF of $X_1,\dots,X_N$, $f_{X_1,\dots,X_N}$, find the joint PDF of $Y_1,\dots,Y_N$, where $Y_1=g_1(X_1,\dots,X_N),\dots,Y_N=g_N(X_1,\dots,X_N)$.

(1) Invert and express X in terms of Y. $X_1=h_1(Y_1,\dots,Y_N),\dots,X_N=h_N(Y_1,\dots,Y_N)$.

(2) Replace $f_{X_1,\dots,X_N}(x_1,\dots,x_N)$ with $f_{X_1,\dots,X_N}(h_1(Y_1,\dots,Y_N),h_2,\dots,h_N)$.

(3) Multiply by the absolute value of the Jacobian $J=\det\left(\frac{\partial X_i}{\partial Y_j}\right)$.

(4) Include the feasible region.

In summary, $f_{Y_1,\dots,Y_N} = f_{X_1,\dots,X_N}(h_1,\dots,h_N)\,|J|$, for feasible region of Y.

Ex: $f_{X_1,X_2} = e^{-(x_1+x_2)}$, $x_1>0$, $x_2>0$, $Y_1 = \frac{X_1}{X_1+X_2}$, $Y_2 = X_1+X_2$. We have $X_1 = Y_1Y_2$ and $X_2 = Y_2 - Y_1Y_2$. Then $J = \det\begin{bmatrix} \partial X_1/\partial Y_1 & \partial X_1/\partial Y_2 \\ \partial X_2/\partial Y_1 & \partial X_2/\partial Y_2 \end{bmatrix} = \det\begin{bmatrix} Y_2 & Y_1 \\ -Y_2 & 1-Y_1 \end{bmatrix} = Y_2$. Thus, $f_{Y_1,Y_2}(y_1,y_2) = f_{X_1X_2}(y_1y_2, y_2-y_1y_2) = e^{-(y_1y_2+(y_2-y_1y_2))}\cdot y_2 = y_2e^{-y_2}$. Our feasible region is $Y_1Y_2>0$ and $Y_2-Y_1Y_2>0$, or $0<Y_1<1$ and $Y_2>0$. Note that Y_1 and Y_2 are independent.

Ex: (x,y) chosen uniformly from inside unit circle. Find joint PDF of the polar coordinates. $R=\sqrt{X^2+Y^2}$, $\Theta = \tan^{-1}(\frac{Y}{X})$. Here, $f_{X,Y}(x,y) = \frac{1}{\pi}$, $0<X^2+Y^2<1$. We have $X = R\cos\Theta$, $Y = R\sin\Theta$, $J = \begin{bmatrix} \cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta \end{bmatrix} = R > 0$. Then $f_{R,\Theta}(r,\theta) = \frac{1}{\pi}r$, $0<r<1$, $0<\theta<2\pi$, and $f_\Theta(\theta) = \frac{1}{2\pi}$, $f_R(r) = 2r$ with the same feasible region.

Ex: Generate a standard normal random variable from a $U(0,1)$ random variable. There is no simple function of X that has $N(0,1)$ PDF since we have no closed form for Φ or Φ^{-1}. However, we can generate 2 independent normals from 2 independent uniform random variables.

Thm: Let X_1, X_2 be iid $U(0,1)$. Define $Y_1 = \sqrt{-2\ln X_1}\,\sin(2\pi X_2)$, $Y_2 = \sqrt{-2\ln X_1}\,\cos(2\pi X_2)$. Then Y_1, Y_2 are iid $N(0,1)$.

Pf: Proof sketch. $Y_1/Y_2 = \tan(2\pi X_2) \Rightarrow X_2 = \frac{\tan^{-1}(Y_1/Y_2)}{2\pi}$, $X_1 = e^{-(Y_1^2+Y_2^2)/2}$, $J = \frac{1}{2\pi}e^{-\frac{1}{2}(y_1^2+y_2^2)}$, $f_{Y_1,Y_2}(y_1,y_2) = 1\cdot|J| = \frac{1}{2\pi}e^{-\frac{1}{2}(y_1^2+y_2^2)} = \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_1^2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_2^2}\right)$.

Def: If $X_1,\dots,X_N$ have joint PDF $f_{X_1,\dots,X_N}$, we define $E[g(X_1,\dots,X_N)] = \int\cdots\int g(x_1,\dots,x_N)f_{X_1,\dots,X_N}(x_1,\dots,x_N)\,dx_1\,dx$ and $E(X_1) = \int\cdots\int x_1 f_{X_1,\dots,X_N}(x_1,\dots,x_N)\,dx_1\dots dx_N$.

Thm: $E(X+Y) = E(X)+E(Y)$.

Pf: For continuous case, $E(X+Y) = \iint (x+y)f_{X,Y}(x,y)\,dx\,dy = \iint x f_{X,Y}(x,y)\,dx\,dy + \iint y f_{X,Y}(x,y)\,dx\,dy = E(X)+E(Y)$.

Note: In general, $E(X_1+\dots+X_N) = E(X_1)+\dots+E(X_N)$.

Ex: Show if $X\sim Bin(n,p)$, then $E(X)=np$. Note X = # of heads when flipping a coin, with head probability p, n times. Then $X = X_1+\dots+X_N$ where $X_i = 1$ if the i^{th} flip is heads and 0 if tails. Notice $E(X_i) = 1\cdot p + 0(1-p) = p$. Thus, $E(X) = np$.

Thm: If X and Y are independent, then $E(XY) = E(X)E(Y)$.

Pf: (Continuous case) Suppose X,Y have joint PDF $f_{X,Y}$. Then $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f_{X,Y}(x,y)\,dx\,dy$
$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f_X(x)f_Y(y)\,dx\,dy = \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\,dy\right) = E(X)E(Y)$.

Note: It is not always true that $Var(X_1 + \ldots + X_N) = Var(X_1) + \ldots + Var(X_N)$.

Def: The covariance of X and Y is $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$ where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.

Note: $Cov(X,X) = Var(X)$

Thm: $Cov(X,Y) = E(XY) - E(X)E(Y)$

Pf: $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y] = E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X \mu_Y$
$= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y)$.

Cor: If X and Y are independent, then $Cov(X,Y) = 0$.

Pf: $E(XY) = E(X)E(Y)$ when X and Y are independent.

Ex:

$X \backslash Y$	0	1
-1	0	1/3
0	1/3	0
1	0	1/3

Here, $E(X) = 0$, $E(Y) = 2/3$, $Cov(X,Y) = E(XY) - E(X)E(Y) = 0 - 0(\frac{2}{3}) = 0$.
Note that X and Y are not independent.

Def: X and Y are uncorrelated when $Cov(X,Y) = 0$. When $Cov(X,Y) > 0$, X and Y are positively correlated. When $Cov(X,Y) < 0$, X and Y are negatively correlated.

Note: Covariance is bilinear, i.e. $Cov(aX, bY) = ab\,Cov(X,Y)$ and $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$.

Thm: $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

Pf: Since $E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$, $Var(X+Y) = E[[(X+Y) - (\mu_X + \mu_Y)]^2] = E[[(X-\mu_X) + (Y-\mu_Y)]^2]$
$= E[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)] = E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] + 2E[(X-\mu_X)(Y-\mu_Y)]$
$= Var(X) + Var(Y) + 2Cov(X,Y)$

Cor: If X and Y are independent, $Var(X+Y) = Var(X) + Var(Y)$, $Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y)$, $Var(X-Y) = Var(X) + Var(Y)$.

Thm: $Var(X_1 + \ldots + X_N) = \sum_{i=1}^{N} \sum_{j=1}^{N} Cov(X_i, X_j)$ for any $X_1, \ldots, X_N$. This can also be written as $Var(X_1 + \ldots + X_N) = \sum_{i=1}^{N} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) = \sum_{i=1}^{N} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$.

Cor: If $X_1, \ldots, X_N$ are independent, then $Var(X_1 + \ldots + X_N) = \sum_{i=1}^{N} Var(X_i)$.

Def: Let X have mean μ_X and variance σ_X^2, and Y have mean μ_Y and variance σ_Y^2. We define the correlation as

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Thm: For any random variables X and Y, $-1 \le \rho(X,Y) \le 1$. Moreover, $\rho(X,Y) = \pm 1$ iff Y is a linear function of X.

Def: For a random variable X, we define its moment generating function (MGF) as $M_X(t) = E(e^{tX})$. For any MGF, $M_X(0) = E[e^{0x}] = 1$.

Note: In general, $M_X^{(n)}(0) = E(X^n)$, the n^{th} moment of X.

Ex: $X \sim Poi(\lambda)$, $M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} P(X=k) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$

Ex: $Z \sim N(0,1)$, $M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} f_Z(z)\, dz = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2tz + t^2 - t^2)} dz = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2} dz = e^{t^2/2}$$

Thm: If $Y = aX + b$ then $M_Y(t) = e^{bt} M_X(at)$.

Pf: $M_Y(t) = E(e^{tY}) = E(e^{t(aX+b)}) = E(e^{tb} e^{taX}) = e^{tb} E(e^{taX}) = e^{tb} M_X(at)$.

Ex: Find $M_X(t)$ for $X \sim N(\mu, \sigma^2)$. Note $X = \sigma Z + \mu$ where $Z \sim N(0,1)$. Then we have $M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{(\sigma t)^2/2}$.

Claim: If X_1 and X_2 have the same MGF, then they have the same distribution, i.e. if $M_{X_1}(t) = M_{X_2}(t)$ then $f_{X_1}(x) = f_{X_2}(x)$.

Thm: If $X_1, \dots, X_N$ are independent where X_i has MGF $M_{X_i}(t)$ then $Y = X_1 + X_2 + \dots + X_N$ has MGF $M_Y(t) = \prod_{i=1}^{N} M_{X_i}(t)$. The MGF of the sum is the product of the MGFs.

Pf: $M_Y(t) = E(e^{tY}) = E(e^{t(X_1 + \dots + X_N)}) = E(e^{tX_1} e^{tX_2} \dots e^{tX_N}) = E(e^{tX_1}) \dots E(e^{tX_N}) = \prod_{i=1}^{N} M_{X_i}(t)$.

Ex: $X \sim Bin(n,p)$. Find $M_X(t)$. Note $X = X_1 + \dots + X_N$ where $X_i = \begin{cases} 1, & \text{probability } p \\ 0, & \text{probability } q \end{cases}$. Note $M_{X_i} = E(e^{tX_i}) = pe^{t \cdot 1} + qe^{t \cdot 0} = pe^t + q$. Thus, $M_X(t) = (pe^t + q)^N$.

Ex: If $X_i \sim Poi(\lambda_i)$ and $X = X_1 + \dots + X_N$ then $X \sim Poi(\lambda_1 + \dots + \lambda_N)$.

Ex: If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Pf: $M_{X_i}(t) = e^{\mu_i t + \frac{1}{2}(\sigma_i t)^2}$

$$\Rightarrow M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t) = e^{\mu_1 t + \frac{1}{2}(\sigma_1 t)^2} e^{\mu_2 t + \frac{1}{2}(\sigma_2 t)^2} = e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}.$$

Thm: (Markov's Inequality) For any nonnegative random variable X and any real $t > 0$,

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Pf: Let X be a nonnegative continuous random variable. For any fixed $t > 0$, suppose X has PDF $f_X(x)$.

$$E(X) = \int_0^{\infty} x f_X(x)dx = \int_0^t x f_X(x)dx + \int_t^{\infty} x f_X(x)dx \geq \int_t^{\infty} x f_X(x)dx \geq \int_t^{\infty} t f_X(x)dx = t\int_t^{\infty} f_X(x)dx = tP(X \geq t).$$

Thm: (Chebyshev's Inequality) If $E(X) = \mu$ and $Var(X) = \sigma^2$ then for any $t > 0$,

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Pf: Let $Y = (X-\mu)^2$. Note $Y \geq 0$ and $E(Y) = E[(X-\mu)^2] = Var(X)$ so by Markov's inequality,

$$P(|X-\mu| \geq t) = P((X-\mu)^2 \geq t^2) = P(Y \geq t^2) \leq \frac{E(Y)}{t^2} = \frac{Var(X)}{t^2}.$$

Note: By Chebyshev's inequality, $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$.

Def: Suppose $X_1, \ldots, X_N$ are i.i.d. random variables each with mean μ and variance σ^2. Their sample mean $\overline{X}_N = (X_1 + \ldots + X_N)/N$ is a random variable.

Thm: $E(\overline{X}_N) = \mu$ and $Var(\overline{X}_N) = \sigma^2/N$.

Pf: $E(\overline{X}_N) = E[(X_1 + \ldots + X_N)/N)] = \frac{1}{N} E(X_1 + \ldots + X_N) = \frac{1}{N}[E(X_1) + \ldots + E(X_N)] = \frac{1}{N}(\mu N) = \mu$.

$Var(\overline{X}_N) = Var[(X_1 +. + X_N)/N)] = \frac{1}{N^2} Var(X_1 + \ldots + X_N) = \frac{1}{N^2}[Var(X_1) + \ldots + Var(X_N)] = \frac{1}{N^2}(\sigma^2 N) = \sigma^2/N$.

Cor: $SD(\overline{X}_N) = \sigma/\sqrt{N}$.

Thm: (Law of Large Numbers) For any $\varepsilon > 0$, $P(|\overline{X}_N - \mu| < \varepsilon) \to 1$ as $N \to \infty$.

Pf: By Chebyshev's inequality, for any $\varepsilon > 0$, $P(|\overline{X}_N - \mu| \geq \varepsilon) \leq \frac{Var(\overline{X}_N)}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2}$. Thus,

$P(|\overline{X}_N - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{N\varepsilon^2}$. As $N \to \infty$, $P(|\overline{X}_N - \mu| < \varepsilon) \to 1$.

Thm: (Central Limit Theorem) The sum or average of a large number of independent random variables has approximately a normal distribution. Suppose $X_1, \ldots, X_N$ are i.i.d. with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ then $X_1 + \ldots + X_N \approx N(N\mu, N\sigma^2)$. Equivalently, $\overline{X}_N \approx N(\mu, \sigma^2/N)$ and $\frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}} \approx N(0,1)$.

Pf: Let $Y = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}}$. We will show as $N \to \infty$, $M_Y(t) \to e^{t^2/2}$, the MGF of $N(0,1)$. For $i = 1, \ldots, N$, let $Y_i = \frac{X_i - \mu}{\sigma}$. Then $E(Y_i) = E\left[\frac{X_i - \mu}{\sigma}\right] = \frac{E(X_i) - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$ and $Var(Y_i) = Var\left(\frac{X_i - \mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X_i) = \sigma^2/\sigma^2 = 1$ and $E(Y_i^2) = Var(Y_i) + (E(Y_i))^2 = 1 + 0^2 = 1$. Note that $\sum_{i=1}^{N} \frac{Y_i}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sum \frac{X_i - \mu}{\sigma} = \frac{\Sigma X_i - N\mu}{\sigma\sqrt{N}} = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}} = Y$. Now suppose each Y_i has MGF $M(t)$. By Taylor's Theorem, Y_i has MGF $M(t) = M(0) + M'(0)t + M''(0)\frac{t^2}{2} + M'''(c)\frac{t^3}{6}$ for some $0 < c < t$. Since $M(0) = 1$, $M'(0) = E(Y_i) = 0$, and $M''(0) = E(Y_i^2) = 1$, we have $M(t) = 1 + \frac{t^2}{2} + \frac{M'''(c)}{6} t^3$. Note $Y_i/\sqrt{N}$ has MGF $M(\frac{t}{\sqrt{N}}) = 1 + (t/\sqrt{N})^2/2 + \frac{M'''(c_N)}{6}(t/\sqrt{N})^3 = (1 + t^2/2 + d_N \cdot t^3/\sqrt{N})/N$ where $0 < c_N < t/\sqrt{N}$ and $d_N = \frac{M'''(c_N)}{6}$. Note $Y = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}} = \Sigma Y_i/\sqrt{N}$ has MGF $(M(\frac{t}{\sqrt{N}}))^N$ so as $n \to \infty$, Y has MGF

$$\lim_{N \to \infty} \left(\frac{1 + t^2/2 + d_N t^3/\sqrt{N}}{N} \right)^N = e^{t^2/2}$$

which is the MGF of $N(0,1)$.