Chapter 5*

September 2, 2018

1 Discrete random variables

1.1 Bernoulli distribution

An experiment, or trial, whose outcome can be classified as either a **success** or **failure** is performed.

Let

$$X = \begin{cases} 1 & \text{when the outcome is a "success"} \\ 0 & \text{when the outcome is a "failure".} \end{cases}$$

If p is the probability of a success then the p.m.f of X is

$$\begin{aligned} p(0) &= \mathbb{P}(X = 0) = 1 - p \\ p(1) &= \mathbb{P}(X = 1) = p \\ p(x) &= \mathbb{P}(X = x) = 0 \qquad \text{for } x \notin \{0, 1\}. \end{aligned}$$

Definition: A random variable is called a *Bernoulli* random variable if it has the above p.m.f for $p \in [0, 1]$. We write $X \sim \text{Ber}(p)$.

Expected value and variance:

$$\begin{aligned} \mathbb{E}(X) &= 0 \cdot (1 - p) + 1 \cdot p = p. \\ Var(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = p - p^2 = p(1 - p). \end{aligned}$$

Example: Flip a fair coin. Let $X = \#$ of heads. Then X is a Bernoulli random variable with $p = \frac{1}{2}$. Thus,

$$\mathbb{E}(X) = \frac{1}{2}, \qquad \text{Var}(X) = \frac{1}{4}.$$

*Notes for Chapter 5 of DeGroot and Schervish adapted from Giovanni Motta's, Bodhisattva Sen's and Martin Lindquists notes for STAT W4109/W4105.

1.2 Binomial random variables

Suppose that n independent Bernoulli trials are performed. Each of these trials has probability p of success and probability $(1-p)$ of failure.

Let $X = \#$ of successes in the n trials. Then

$$
\begin{aligned}
p(0) &= \mathbb{P}(0 \text{ successes in } n \text{trials}) = (1-p)^n && \{FFFFFFF\} \\
p(1) &= \mathbb{P}(1 \text{ successes in } n \text{trials}) = \binom{n}{1} p(1-p)^{n-1} && \{FSFFFFF\} \\
p(2) &= \mathbb{P}(2 \text{ successes in } n \text{trials}) = \binom{n}{2} p^2(1-p)^{n-2} && \{FSFSFFF\} \\
&\cdots \\
p(k) &= \mathbb{P}(k \text{ successes in } n \text{trials}) = \binom{n}{k} p^k(1-p)^{n-k} \\
&\cdots
\end{aligned}
$$

Definition: A r.v is said to have a *Binomial* distribution with parameters (n, p) if it has the p.m.f

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \text{for } k = 0, 1, 2, \ldots, n.$$

We usually write $X \sim \mathrm{Bin}(n, p)$.

Note that this is a valid p.m.f: $\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+(1-p))^n = 1$.

Example: A Ber(p) random variable is Bin$(1, p)$.

Theorem 1.1. *Suppose that $X_1, \ldots, X_n$ are i.i.d (independent and identically distributed) Bernoulli r.v's with parameter p. Then*

$$X = X_1 + \ldots + X_n \sim \mathrm{Bin}(n, p).$$

Example: Roll a die 3 times. Find the p.m.f of the number of times we roll a 5.

Solution: Let $X = \#$ of times we roll a 5 (number of 'successes'). Then $X \sim$

Binomial($3, \frac{1}{6}$). Thus,

$$\begin{aligned}
p(0) &= \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = \frac{125}{216}, \\
p(1) &= \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = \frac{75}{216}, \\
p(2) &= \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = \frac{15}{216}, \\
p(3) &= \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 = \frac{1}{216},
\end{aligned}$$

Example: Screws produced by a certain company will be defective with probability 0.01 independently of each other. If the screws are sold in packages of 10, what is the probability that two or more screws are defective?

Solution: Let $X = \#$ of defective screws. Then, $X \sim$ Binomial(10,0.01). Thus,

$$\begin{aligned}
\mathbb{P}(X \geq 2) &= 1 - \mathbb{P}(X < 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \\
&= 1 - \binom{10}{0}(.01)^0(0.99)^{10} - \binom{10}{1}(.01)^1(0.99)^9 \approx .004.
\end{aligned}$$

Expected value of a Bin(n, p) r.v:

$$\mathbb{E}(X) = np \qquad \text{(use linearity of expectation).}$$

Variance of a Bin(n, p) r.v:

$$\mathrm{Var}(X) = np(1-p) \qquad \text{(Use rule for adding variances of i.i.d r.v's).}$$

1.2.1 Properties of a Binomial random variable

Example: Sums of two independent Binomial random variables.

Suppose that $X \sim \mathrm{Bin}(n, p)$ and $Y \sim \mathrm{Bin}(m, p)$, and that X and Y are independent. Let $Z = X + Y$. Find the distribution of Z.

Use convolution formula: for $0 \leq z \leq n + m$,

$$\begin{aligned}
p_Z(z) &= \sum_{k=0}^{n} p_X(k) p_Y(z-k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \binom{m}{z-k} p^{z-k}(1-p)^{m-z+k} \\
&= p^z(1-p)^{n+m-z} \sum_{k=0}^{n} \binom{n}{k} \cdot \binom{m}{z-k} = \binom{n+m}{z} p^z (1-p)^{n+m-z}.
\end{aligned}$$

Thus, $Z \sim \text{Bin}(n+m, p)$.

Can also be proved using m.g.f's:

$$\begin{aligned} M_X(t) &= (pe^t + 1 - p)^n \\ M_Y(t) &= (pe^t + 1 - p)^m. \end{aligned}$$

Now use the uniqueness theorem for m.g.f's.

1.3 Multinomial random variable

In the binomial case, we counted the number of "successes" in a binary experiment (e.g., the number of heads in a coin toss experiment). The multinomial case generalizes this from the binary to the k-ary case.

Let, for $k \geq 0$,

$$\begin{aligned} N_i &= \#\text{ of observed outcomes in } i\text{-th slot}, \quad \text{and} \\ p_i &= \text{probability of falling in } i\text{-th slot.} \end{aligned}$$

Then, for $n_i \geq 0$ and such that $n_1 + n_2 + \ldots + n_k = n$,

$$p(n_1, n_2, \ldots, n_k) = \binom{n}{n_1, n_2, \ldots, n_k} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k} = \frac{n!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}.$$

The first term is the usual multinomial combinatorial term (familiar from Chapter 1); the second is the probability of observing any sequence with n_1 1's, n_2 2's, ..., and n_k k's.

To compute the moments here, just note that $N_i \sim \text{Bin}(n, p_i)$. So,

$$\mathbb{E}(N_i) = np_i, \qquad \text{and} \qquad \text{Var}(N_i) = np_i(1 - p_i).$$

For the covariances, note that $N_i + N_j \sim \text{Bin}(n, p_i + p_j)$. So

$$\text{Var}(N_i + N_j) = n(p_i + p_j)(1 - p_i - p_j).$$

But we also know that

$$\text{Var}(N_i + N_j) = \text{Var}(N_i) + \text{Var}(N_j) + 2\text{Cov}(N_i, N_j).$$

Put these together we get that

$$\text{Cov}(N_i, N_j) = -np_i p_j.$$

Note that it makes sense that the covariance is negative: since n is limited, the more of N_i we see, the fewer N_j's we'll tend to see.

1.4 The Poisson random variable

A r.v X, taking on the values $0, 1, 2, \ldots$, is said to have a *Poisson* distribution with parameter λ ($\lambda > 0$) if

$$p(i) = \mathbb{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!} \qquad \text{for } i = 0, 1, 2, 3, \ldots.$$

This is a valid p.m.f:

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1.$$

The Poisson random variable has a large range of applications – examples include number of arrivals of customers for service, number of phone calls, number of occurrences of floods, etc. in a *fixed time period*.

A major reason for this is that a Poisson random variable can be used as an approximation for a Binomial random variable with parameter (n, p) when n is large and p is small (rare events).

Let $X \sim \mathrm{Bin}(n, p)$ and let $\lambda = np$ (λ moderate). Then,

$$\begin{aligned}
\mathbb{P}(X = i) &= \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!i!}p^i(1-p)^{n-i} \\
&= \frac{n(n-1)\ldots(n-i+1)}{i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} = \frac{n(n-1)\ldots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}
\end{aligned}$$

For n large and i, λ moderate,

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}; \qquad \frac{n(n-1)\ldots(n-i+1)}{n^i} \approx 1; \qquad \left(1-\frac{\lambda}{n}\right)^i \approx 1.$$

Hence,

$$\mathbb{P}(X = i) \approx e^{-\lambda}\frac{\lambda^i}{i!}.$$

Result: If X and Y are independent Poisson random variables with parameters λ_1 and λ_2 respectively, then $Z = X + Y \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$.

Solution: Verify for yourself, using either m.g.f (easiest) or convolution formula.

1.4.1 Rare event

Example: A typesetter, on the average makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?

Solution: Assume that each word is a Bernoulli trial with probability of success 1/500 and that the trials are independent.

Let X = # of errors in five pages (1500 words). Thus, $X \sim \text{Binomial}(1500, \frac{1}{500})$, and

$$\mathbb{P}(X \leq 2) = \sum_{x=0}^{2} \binom{1500}{x} \left(\frac{1}{500}\right)^x \left(\frac{499}{500}\right)^{1500-x} \approx 0.4230.$$

Use the Poisson approximation with $\lambda = np = 1500/500 = 3$. Then $X \sim \text{Poi}(3)$ (approximately) and

$$\mathbb{P}(X \leq 2) \approx e^{-3} + 3e^{-3} + \frac{3^2 e^{-3}}{2} \approx 0.4232.$$

Result: If $X \sim \text{Poi}(\lambda)$, then $\mathbb{E}(X) = \lambda$ and $\text{Var}(X) = \lambda$.

Observe that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda}\lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda \\
\mathbb{E}(X(X-1)) &= \sum_{i=0}^{\infty} i(i-1) e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda}\lambda^2 \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} = e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2.
\end{aligned}$$

Thus,

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}(X) - [\mathbb{E}(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

(Compare with Binomial: $np(1-p)$, n large, p small $\Rightarrow$ variance $\approx np$)

Result: The moment generating function of X is given by

$$\mathbb{E}(e^{tX}) = \sum_{i=0}^{\infty} e^{it} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

Note that this is the limit of the binomial case: $(1 + (e^t - 1)p)^n$.

1.5 Geometric distribution

Suppose that independent Bernoulli trials are being performed. Each of these trials has probability p of success and probability $(1-p)$ of failure.

Let X = # of unsuccessful trials preceding the first success.

Then X is a discrete random variable that can take on values in $\{0, 1, 2, 3, \ldots\}$.

Possible outcomes: $S, FS, FFS, FFFS, FFFFS, FFFFFS, FFFFFFS, \ldots$.

$$\begin{array}{rcl} \mathbb{P}(X = 0) & = & p \\ \mathbb{P}(X = 1) & = & (1 - p)p \\ \mathbb{P}(X = 2) & = & (1 - p)^2 p \\ & \cdots & \\ \mathbb{P}(X = n) & = & (1 - p)^n p \\ & \cdots & \end{array}$$

A r.v is said to have a *Geometric* distribution with parameter p if it has the p.m.f

$$p(k) = (1 - p)^k p \qquad\qquad \text{for } k = 0, 1, 2, \ldots.$$

This is a valid p.m.f:

$$\sum_{k=0}^{\infty} p(1 - p)^k = \frac{p}{1 - (1 - p)} = 1.$$

Show that for the geometric distribution, $\mathbb{P}(X \geq a) = (1 - p)^a$, for $a = 0, 1, \ldots$.

Example: A fair coin is flipped until a head appears. What is the probability that the coin needs to be flipped more than 5 times?

Solution: Let $X = \#$ of unsuccessful trials preceding the first H.

$$\mathbb{P}(X \geq 5) = (1 - p)^5 = \frac{1}{32}.$$

This is the same probability as getting 5 straight tails. (Note that the combinatoric term here is $\binom{5}{5} = 1$.)

Expected value: Write $q = 1 - p$. Then,

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} n q^n p = pq \sum_{n=0}^{\infty} \frac{d}{dq}(q^n) = pq \frac{d}{dq}\left(\frac{1}{1 - q}\right) = \frac{pq}{(1 - q)^2} = \frac{q}{p}.$$

We have used the following result (on converging power series)

$$\frac{d}{dq} \sum_{n=0}^{\infty} a_n q^n = \sum_{n=0}^{\infty} \frac{d}{dq}(a_n q^n).$$

Variance: Use $\mathrm{Var}(X) = \mathbb{E}(X^2) - [E(X)]^2$, and use m.g.f to get $\mathbb{E}(X^2)$. Note that

$$M_X(t) = \sum_{n=0}^{\infty} e^{tn} q^n p = \sum_{n=0}^{\infty} (qe^t)^n p = \frac{p}{1 - qe^t},$$

for $qe^t < 1$, otherwise infinite. Take second derivative and show that $\mathrm{Var}(X) = \frac{q}{p^2}$.

1.6 Negative Binomial

As before, suppose that independent Bernoulli trials are being performed. Each of these trials has probability p of success and probability $(1-p)$ of failure.

Let $X = \#$ of *failures that occur before* the r-th success.

Then X takes values in $\{0, 1, 2, \ldots\}$ and

$$\begin{aligned}
\mathbb{P}(X=0) &= p^r && \text{SSSSSSS} \\
\mathbb{P}(X=1) &= \binom{r}{1} p^r (1-p) && \text{SSSSFSSS} \\
\mathbb{P}(X=2) &= \binom{r+1}{2} p^r (1-p)^2 && \text{SSSSFSFSS} \\
&\cdots \\
\mathbb{P}(X=n) &= \binom{r+n-1}{n} p^r (1-p)^n
\end{aligned}$$

Definition: A r.v has a *negative binomial* distribution with parameters (r, p), $r = 1, 2, \ldots$ and $p \in (0,1)$, if it has the p.m.f

$$p(n) = \binom{r+n-1}{n} p^r (1-p)^n, \qquad \text{for } n = 0, 1, 2, \ldots.$$

This is a valid p.d.f: Need to use negative binomials.

The negative binomial distribution is used when the number of successes is fixed and we're interested in the number of failures before reaching the fixed number of successes.

If $X \sim \text{NegBin}(r, p)$, then

$$X = \sum_{i=1}^{r} X_i,$$

with X_i are independent $\text{Geo}(p)$ r.v's. Thus,

$$\mathbb{E}(X) = \frac{rq}{p}, \qquad \text{Var}(X) = \frac{rq}{p^2}.$$

MGF of $\text{Geo}(p)$:

$$\mathbb{E}(e^{tX_i}) = \sum_{n=0}^{\infty} e^{tn} q^n p = p \sum_{n=0}^{\infty} (qe^t)^n = \left[\frac{p}{1-qe^t}\right]$$

for $qe^t < 1$.

Thus, MGF of $\text{NegBin}(r, p)$ is $\left[\frac{p}{1-qe^t}\right]^r$.

Example: Find the expected number of times one needs to roll a die before getting 4 sixes.

Solution: Let $X = \#$ of rolls before getting 4 sixes. Then, $X = Y + 3$, where Y has a negative binomial distribution with $r = 4$ and $p = 1/6$. Then

$$\mathbb{E}(Y) = r\frac{q}{p} = 4 \cdot \frac{5/6}{1/6} = 20,$$

and thus, $\mathbb{E}(X) = 23$.

1.7 Hypergeometric random variables

Given a population of size N with two types of objects, m being of **type 1** and $N - m$ of **type 2**, we are interested in the number of objects of type 1 in a sample of n extracted objects.

Example: Assume we have a box that contains m red balls and $(N - m)$ white balls. Suppose we choose n different balls from the box, without replacement. What is the distribution of the number of red balls drawn?

Solution: Let $X = \#$ of red balls. Then, X takes values from 0 to n (assuming $N - m \geq n$ and $m \geq n$).

$$\begin{aligned}
\mathbb{P}(X = 0) &= \frac{\binom{m}{0}\binom{N-m}{n}}{\binom{N}{n}} \\
\mathbb{P}(X = 1) &= \frac{\binom{m}{1}\binom{N-m}{n-1}}{\binom{N}{n}} \\
&\cdots \\
\mathbb{P}(X = k) &= \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \\
&\cdots
\end{aligned}$$

Definition: A r.v is said to have a *Hypergeometric* distribution with parameters (n, N, m) if it has the p.m.f

$$\mathbb{P}(X = k) = \begin{cases} \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} & \text{for } \max\{0, n - (N - m)\} \leq k \leq \min\{n, m\} \\ 0 & \text{otherwise.} \end{cases}$$

Exercise: Show that this is a valid p.m.f?

Suppose that we have a population of size N that consists of two types of objects. For example, we could have balls in an urn that are red or green, a population of people who are either male or female, etc. Assume there are m objects of type 1, and $N - m$ objects of type 2.

Let $X = \#$ of objects of type 1 in a sample of n objects.

Then $X \sim \text{HyGeo}(n, N, m)$.

Example: Suppose 25 screws are made, of which 6 are defective. Suppose we randomly sample 10 screws and place them in a package. What is the probability that the package contains no defective screws?

Solution: Let $X = \#$ of defective screws in the package.

Then X has a hypergeometric distribution with $n = 10, N = 25$ and $m = 6$ and thus,

$$\mathbb{P}(X = 0) = \frac{\binom{6}{0}\binom{19}{10}}{\binom{25}{10}} \approx 0.028.$$

Expected value: Using $x\binom{m}{x} = m\binom{m-1}{x-1}$ and $\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1}$, we find that

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x=0}^{n} x \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} = \sum_{x=0}^{n} \frac{m\binom{m-1}{x-1}\binom{N-m}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} \\ &= \frac{mn}{N} \sum_{x=0}^{n} \frac{\binom{m-1}{x-1}\binom{N-m}{n-x}}{\binom{N-1}{n-1}} = \frac{mn}{N}.\end{aligned}$$

Variance:

$$\text{Var}(X) = \frac{mn}{N} \frac{(N-m)(N-n)}{N(N-1)}.$$

1.7.1 Sampling without replacement

Suppose that a box contains A red balls and B blue balls. Suppose that $n \geq 1$ balls are selected at random from the box *without replacement*, and let X denote the number of red balls obtained.

Then $X \sim \text{HyGeo}(n, A+B, A)$.

Question: What is $\mathbb{E}(X)$ and $\text{Var}(X)$?

Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th drawn ball is red,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Then, each X_i is a Bernoulli r.v, and

$$X = X_1 + \ldots + X_n.$$

But $X_1, X_2, \ldots, X_n$ are **not independent** (Why?). Show that

$$\mathbb{P}(X_2 = 1 | X_1 = 0) = \frac{A}{A+B-1} \neq \frac{A-1}{A+B-1} = \mathbb{P}(X_2 = 1 | X_1 = 1).$$

However, note that $\mathbb{P}(X_i = 1) = \frac{A}{A+B}$ for all $i = 1, \ldots, n$ (Prove this. Hint: We can imagine that the n balls are selected from the box by first arranging all the balls in the box in some random order and then selecting the first n balls from the arrangement).

This shows that

$$\mathbb{E}(X_i) = \frac{A}{A+B} \qquad \text{and} \qquad \mathrm{Var}(X_i) = \frac{AB}{(A+B)^2}.$$

Thus,

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{nA}{A+B}.$$

To find $\mathrm{Var}(X)$ observe that

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j) = n\frac{AB}{(A+B)^2} + n(n-1)\mathrm{Cov}(X_1, X_2)$$

by symmetry among the r.v's $X_1, \ldots, X_n$.

Note that

$$X_1 X_2 = \begin{cases} 1 & \text{if both the 1-st and 2-nd drawn balls are red,} \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$\begin{aligned} \mathbb{E}(X_1 X_2) &= \mathbb{P}(X_1 X_2 = 1) = \mathbb{P}(X_2 = 1, X_1 = 1) \\ &= \mathbb{P}(X_2 = 1 | X_1 = 1)\mathbb{P}(X_1 = 1) = \frac{A-1}{A+B-1}\frac{A}{A+B}. \end{aligned}$$

This gives

$$\mathrm{Cov}(X_1, X_2) = \frac{A-1}{A+B-1}\frac{A}{A+B} - \left(\frac{A}{A+B}\right)^2 = -\frac{AB}{(A+B-1)(A+B)^2}$$

after some simplification. We thus have

$$\mathrm{Var}(X) = n\frac{AB}{(A+B)^2} - \frac{n(n-1)AB}{(A+B-1)(A+B)^2} = n\frac{AB}{(A+B)^2}\frac{A+B-n}{(A+B-1)}.$$

1.8 Sampling with replacement

Suppose that a box contains A red balls and B blue balls. Suppose that $n \geq 1$ balls are selected at random from the box *with replacement*, and let X denote the number of red balls obtained.

Here the X_i's, defined in (1), are *independent* and thus $X \sim \text{Bin}(n, A/(A+B))$.

As population size N becomes large, hypergeometric probabilities converge to binomial probabilities (sampling without replacement $\to$ sampling with replacement).

To see this, just set $m = Np$ (where $p =$ fraction of red balls), and write out

$$\begin{aligned}
\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} &= \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}} = \frac{(Np)!}{k!(Np-k)!} \times \frac{[N(1-p)]!}{(n-k)!(N-Np-n+k)!} \times \frac{n!(N-n)!}{N!} \\
&= \binom{n}{k} \frac{[(Np)(Np-1)\cdots(Np-k+1)]}{N(N-1)\cdots(N-n+1)} \times \\
&\quad [(N(1-p))(N(1-p)-1)\cdots(N(1-p)-n+k+1)] \\
&\approx \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

2 Continuous random variables

2.1 Uniform random variables

Definition: X is said to be a *uniform* random variable on (α, β) if it has a p.d.f given by

$$f(u) = \begin{cases} \frac{1}{\beta-\alpha} & \text{for } \alpha < u < \beta, \\ 0 & \text{otherwise.} \end{cases}$$

Expected value and variance: Show that

$$\mathbb{E}(X) = \frac{\alpha+\beta}{2}, \qquad \text{and} \qquad \text{Var}(X) = \frac{(\beta-\alpha)^2}{12}.$$

Example: Buses arrive at specified stop at 15-minute intervals starting at 7 AM (7:00, 7:15, 7:30, 7:45, ...). If the passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30 AM, find the probability that he waits

(a) less than 5 minutes for the bus,

(b) more than 10 minutes for the bus.

Solution: Let $X = \#$ of minutes past 7 AM the passenger arrives.

Then, X is a uniform random variable over the interval $(0, 30)$.

(a) The passenger waits less than 5 minutes if he arrives either between 7:10 and 7:15 or between 7:25 and 7:30. Thus the required probability is $\frac{5}{30} + \frac{5}{30} = \frac{1}{3}$.

(b) The passenger waits more than 10 minutes if he arrives either between 7:00 and 7:05 or between 7:15 and 7:20. Thus the required probability is again $\frac{1}{3}$.

2.2 Exponential random variables

Definition: X is an *exponential* random variable with parameter λ ($\lambda > 0$) if the p.d.f of X is given by

$$f(v) = \begin{cases} \lambda e^{-\lambda v} & \text{for } v \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The c.d.f of X is given by

$$F(a) = \mathbb{P}(X \leq a) = \int_0^a \lambda e^{-\lambda v} dv = 1 - e^{-\lambda a}, \qquad \text{for } a \geq 0.$$

The exponential distribution can be used to model lifetimes or the time between unlikely events. They are used to model waiting times, at telephone booths, at the post office and for time until decay of an atom in radioactive decay.

Expected value and variance: Show that

$$\mathbb{E}(X) = \frac{1}{\lambda} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Show that $\mathbb{E}(X^2) = \frac{2}{\lambda^2}$. The standard deviation (the "scale" of the density) is $\frac{1}{\lambda}$.

These formulas are easy to remember if you keep in mind the following fact: if $X \sim \mathrm{Exp}(1)$ and $Y = X/\lambda, \lambda > 0$, then $Y \sim \mathrm{Exp}(\lambda)$. Thus λ is really just a scaling factor for Y.

Example: Suppose that the length of a phone call in minutes is an exponential random variable with $\lambda = 1/10$. If somebody arrives immediately ahead of you at a public telephone booth, find the probability you have to wait

(a) more than 10 minutes;

(b) between 10 and 20 minutes.

Solution: Let X = length of call made by person in the booth.

$$(a) \qquad \mathbb{P}(X > 10) = \int_{10}^{\infty} f(x)dx = e^{-1} \approx 0.368.$$
$$(b) \qquad \mathbb{P}(10 < X < 20) = F(20) - F(10) = e^{-1} - e^{-2} \approx 0.233.$$

Lemma 2.1. *Suppose that $X_1, \ldots, X_n$ are i.i.d Exp(β). Then the distribution of $Y_1 := \min\{X_1, \ldots, X_n\} \sim$ Exp($n\beta$).*

Proof. For any $t > 0$,

$$\begin{aligned}\mathbb{P}(Y_1 > t) &= \mathbb{P}(X_1 > t, \ldots, X_n > t) \\ &= \mathbb{P}(X_1 > t) \ldots \mathbb{P}(X_n > t) \\ &= e^{-\beta t} \cdots e^{-\beta t} = e^{-n\beta t}.\end{aligned}$$

Now comparing this result with the distribution function of an exponential r.v yields the desired result. □

2.2.1 The Memoryless property of the Exponential distribution

Definition: A nonnegative random variable is *memoryless* if

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$$

for all $s, t > 0$.

Note that

$$\mathbb{P}(X > s + t | X > s) = \frac{1 - F(s+t)}{1 - F(s)}.$$

Defining $G(x) = 1 - F(x)$, the memoryless property may be restated as

$$G(s + t) = G(s)G(t)$$

for all $s, t > 0$.

We know that $G(x)$ must be non-increasing as a function of x, and $0 \leq G(x) \leq 1$. It is known that the only such continuous function is

$$G(x) = \begin{cases} 1 & x < 0 \\ e^{-\lambda x}, & x \geq 0, \end{cases}$$

which corresponds to the exponential distribution with rate λ.

Thus we have proved:

Fact 1: Exponential random variables are memoryless.

Fact 2: The Exponential distribution is the only continuous distribution that is memoryless.

There is a similar definition of the memoryless property for discrete r.v's; in this case, the geometric distribution is the only discrete distribution that is memoryless (the proof is similar).

Example: Suppose the lifetime of a light bulb is exponential with $\lambda = 1/1000$. If the light survives 500 hours what is the probability it will last another 1000 hours.

Solution: Let X = the lifetime of the bulb. Then X is exponential with $\lambda = 1/1000$. Thus,

$$\mathbb{P}(X > 1500 | X > 500) = \mathbb{P}(X > 1000) = e^{-1}.$$

2.3 The Gamma random variable

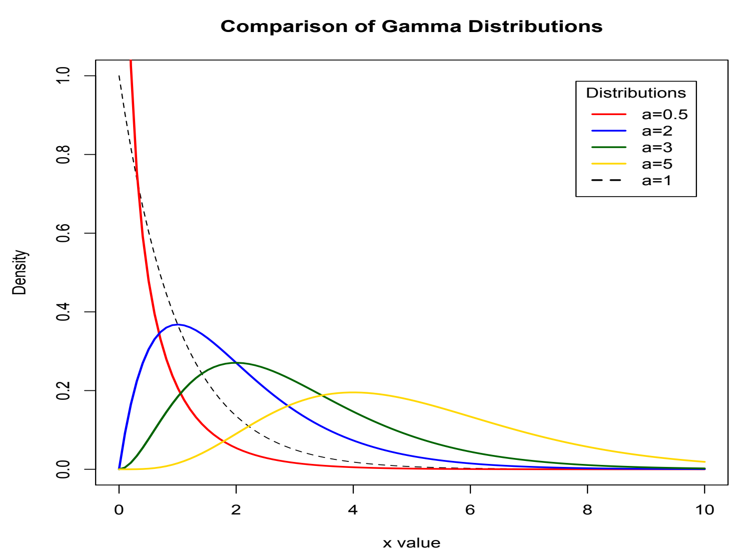
Defintion: X is a Gamma random variable with parameters (α, λ) $(\alpha, \lambda > 0)$ if the p.d.f of X is given by

$$f(x) = \begin{cases} \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0, \end{cases}$$

where

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

Here $\Gamma : (0, \infty) \to \mathbb{R}$ is called the *gamma* function. [Try to plot the p.d.f!] Usually α is called the *shape parameter* and λ is called the *scale parameter*.



It can be shown, for $\alpha > 1$, through integration by parts, that

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1).$$

To see this observe that for $\alpha > 1$, we have

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy = \left[(-1)e^{-y} y^{\alpha-1}\right]_0^\infty - (\alpha-1)\int_0^\infty (-1)e^{-y} y^{\alpha-2} dy = (\alpha-1)\Gamma(\alpha-1).$$

Also a simple calculation shows that $\Gamma(1) = 1$.

Combining these two facts gives us for integer-valued $\alpha = n$,

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\cdots 2\Gamma(1) = (n-1)!.$$

Additionally it can be shown that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. (We will derive it later in the class)

Expected value and variance:

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\Gamma(\alpha)} \int_0^\infty x \cdot \lambda^\alpha e^{-\lambda x} x^{\alpha-1} dx \\
&= \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} \int_0^\infty \frac{1}{\Gamma(\alpha+1)} \lambda^{\alpha+1} e^{-\lambda x} x^{(\alpha+1)-1} dx = \frac{\alpha}{\lambda} \\
\mathbb{E}(X^2) &= \frac{1}{\Gamma(\alpha)} \int_0^\infty x^2 \cdot \lambda^\alpha e^{-\lambda x} x^{\alpha-1} dx \\
&= \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} \int_0^\infty \frac{1}{\Gamma(\alpha+2)} \lambda^{\alpha+2} e^{-\lambda x} x^{(\alpha+2)-1} dx = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \frac{(\alpha+1)\alpha}{\lambda^2} \\
\mathrm{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{(\alpha+1)\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.
\end{aligned}$$

Many random variables can be seen as special cases of the gamma random variable.

(a) If $X \sim \text{Gamma}(\alpha = 1, \lambda)$, then X is an exponential random variable.

(b) If $X \sim \text{Gamma}(\alpha = \frac{n}{2}, \lambda = \frac{1}{2})$ then X has a χ^2 distribution with n degrees of freedom. (We'll discuss this soon)

Example: Let $X \sim \text{Gamma}(\alpha, \lambda)$. Calculate $M_X(t)$.

Solution: Notice that

$$\begin{aligned}
M_X(t) &= \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)} dx = \int_0^\infty \frac{\lambda^\alpha e^{-(\lambda-t)x} x^{\alpha-1}}{\Gamma(\alpha)} dx \\
&= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^\infty \frac{(\lambda-t)^\alpha e^{-(\lambda-t)x} x^{\alpha-1}}{\Gamma(\alpha)} dx = \left(\frac{\lambda}{\lambda-t}\right)^\alpha.
\end{aligned}$$

Thus,

$$M_X'(t) = \alpha \left(\frac{\lambda}{\lambda - t}\right)^{\alpha-1} \frac{\lambda}{(\lambda-t)^2} = \frac{\alpha}{\lambda}\left(\frac{\lambda}{\lambda - t}\right)^{\alpha+1}.$$

Therefore,

$$\mathbb{E}(X) = M_X'(0) = \frac{\alpha}{\lambda}.$$

Note that

$$M_X''(t) = \frac{\alpha(\alpha+1)}{\lambda^2}\left(\frac{\lambda}{\lambda - t}\right)^{\alpha+3},$$

which yields

$$\mathbb{E}(X^2) = M_X''(0) = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

Let X and Y be independent gamma random variables with parameters (s, λ) and (t, λ). Calculate the distribution of $Z = X + Y$.

Thus, $Z = X + Y \sim \text{Gamma}(s + t, \lambda)$. (Can also use the convolution formula)

In general, we have the following proposition.

Proposition 2.2. *If $X_i, i = 1, \ldots, n$, are independent gamma random variables with parameters (t_i, λ) respectively, then $\sum_{i=1}^n X_i \sim$ Gamma$(\sum_{i=1}^n t_i, \lambda)$.*

Proof. Use the example above and prove by induction. □

Example: Sums of independent exponential random variables. Exponential with parameter λ is equivalent to Gamma$(1, \lambda)$. It follows that if $X_i, i = 1, \ldots, n$ are independent exponential random variables with parameter λ, then $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$.

2.4 The Beta random variable

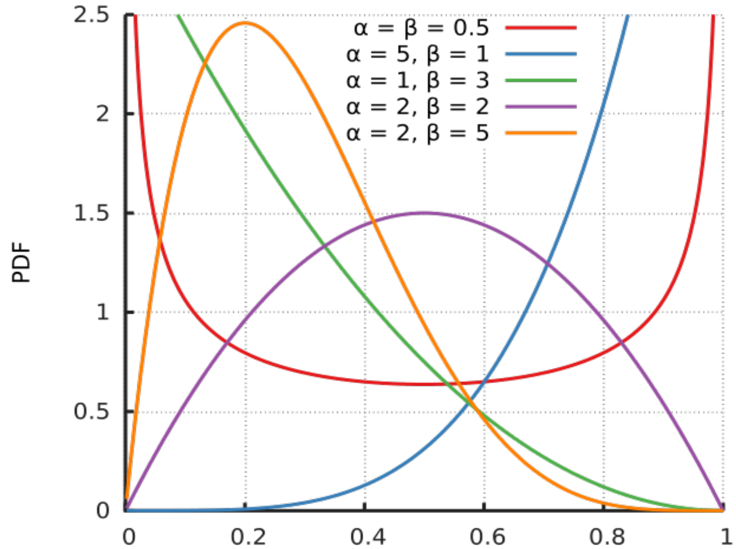
X is a Beta random variable with parameters (a, b) if the p.d.f of X is given by

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & \text{for } x \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

where

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

is called the *Beta function* (Plot the p.d.f!).



The Beta family is a flexible way to model random variables on the interval $[0, 1]$. It is often used to model proportions (e.g., the prior distribution of the fairness of a coin).

The following relationship exists between the beta and gamma function:

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

Expected value and variance:

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{B(a,b)} \int_0^1 x \cdot x^{a-1}(1-x)^{b-1} dx = \frac{B(a+1,b)}{B(a,b)} = \frac{a}{a+b} \\ \mathrm{Var}(X) &= \frac{ab}{(a+b)^2(a+b+1)}. \end{aligned}$$

We will see an application of the beta distribution to order statistics of samples drawn from the uniform distribution soon.

2.5 Normal random variables

Definition: X is a *normal* random variable with parameters μ and σ^2 if the p.d.f of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}. \tag{2}$$

This is the density function for the so-called "bell curve". The normal distribution shows up frequently in probability and statistics, due in large part to the *Central Limit*

Theorem. We write $X \sim N(\mu, \sigma^2)$ when X is normally distributed with parameters μ and σ^2. Normal r.v's are also called Gaussian r.v's.

We can show that f is a valid p.d.f, because $f(x) > 0$ for all $x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f(x)dx = 1$. This last result is not immediately clear.

Proof. The idea is to compute the joint p.d.f of two i.i.d Gaussians, for which the normalization factor turns out to be easier to compute.

If we let $y = (x - \mu)/\sigma$, then

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Let

$$I = \int_{-\infty}^{\infty} e^{-y^2/2} dy.$$

We want to show that $I = \sqrt{2\pi}$.

$$\begin{aligned} I^2 &= I \cdot I = \int_{-\infty}^{\infty} e^{-y^2/2} dy \cdot \int_{-\infty}^{\infty} e^{-z^2/2} dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y^2+z^2)/2} dy\ dz \\ &= \int_{0}^{2\pi} \int_{0}^{\infty} e^{-r^2/2} r dr\ d\theta = 2\pi \int_{0}^{\infty} e^{-v} dv = 2\pi. \end{aligned}$$

where we make the change of variables $y = r\cos\theta$ and $z = r\sin\theta$ (note that $r^2 = y^2 + z^2$). □

Theorem 2.3. *The m.g.f of X, whose p.d.f is given by (2), is*

$$M_X(t) = \mathbb{E}(e^{tX}) = e^{\mu t + \sigma^2 t^2/2}.$$

Proof. By the definition of m.g.f:

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx.$$

By completing the square inside the brackets, we obtain the relationship

$$tx - \frac{(x-\mu)^2}{2\sigma^2} = \mu t + \frac{1}{2}\sigma^2 t^2 - \frac{[x - (\mu + \sigma^2 t)]^2}{2\sigma^2}.$$

Therefore,

$$M_X(t) = C \cdot e^{\mu t + \sigma^2 t^2/2},$$

where

$$C = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{[x-(\mu+\sigma^2 t)]^2}{2\sigma^2}} dx = 1.$$

□

Theorem 2.4. *If $X \sim N(\mu, \sigma^2)$, then $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.*

Proof. The first two derivatives of the m.g.f $M_X(\cdot)$ are

$$\begin{aligned} M'_X(t) &= (\mu + \sigma^2 t)e^{\mu t + \sigma^2 t^2/2} \\ M''_X(t) &= [(\mu + \sigma^2 t)^2 + \sigma^2]e^{\mu t + \sigma^2 t^2/2}. \end{aligned}$$

Plugging $t = 0$ into each of these derivatives yields

$$\mathbb{E}(X) = M'_X(0) = \mu, \qquad \text{and} \qquad \mathrm{Var}(X) = M''_X(0) - [M'_X(0)]^2 = \sigma^2.$$

□

Lemma 2.5. *If $X \sim N(\mu, \sigma^2)$, i.e., X is normally distributed with parameters μ and σ^2, then*

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Proof. For $a > 0$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(aX + b \le y) = \mathbb{P}\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right) \\ f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{1}{a}f_X\left(\frac{y-b}{a}\right), \end{aligned}$$

which is the p.d.f of a normally distributed r.v with parameters $a\mu + b$ and $a^2\sigma^2$. □

2.5.1 Standard normal distribution

If X is a normally distributed random variable with $\mu = 0$ and $\sigma = 1$, then X has a *standard normal* distribution. The p.d.f of X is then given by

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

An important consequence of Lemma 2.5 is that if $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma}$$

is standard normal.

Expected value and variance: Let Z be a standard normal random variable.

$$\mathbb{E}(Z) = \int_{-\infty}^{\infty} z f_Z(z)dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z e^{-z^2/2}dz = 0.$$

We can check that $\text{Var}(Z) = 1$ using the m.g.f.

Let $X \sim N(\mu, \sigma^2)$. Let $Z = \frac{X-\mu}{\sigma}$. Then, $Z \sim N(0,1)$. Thus,

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mu + \sigma Z) = \mu + \sigma\mathbb{E}(Z) = \mu \\ \text{Var}(X) &= \text{Var}(\mu + \sigma Z) = \sigma^2\text{Var}(Z) = \sigma^2.\end{aligned}$$

It is standard to denote the c.d.f of a standard normal random variable by $\Phi(\cdot)$, i.e.,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-v^2/2}dv.$$

The values of $\Phi(z)$ for non-negative z are given in tables.

For negative values of z use the following relationship:

$$\Phi(-z) = \mathbb{P}(Z \leq -z) = \mathbb{P}(Z > z) = 1 - \mathbb{P}(Z \leq z) = 1 - \Phi(z).$$

Also if $X \sim N(\mu, \sigma^2)$ then,

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Example: If X is a normally distributed with parameters $\mu = 3$ and $\sigma^2 = 9$, find

(a) $\mathbb{P}(2 < X < 5)$

(b) $\mathbb{P}(X > 0)$

(c) $\mathbb{P}(|X - 3| > 6)$.

Solution:

(a) $\mathbb{P}(2 < X < 5) = \mathbb{P}(-\frac{1}{3} < Z < \frac{2}{3}) = \Phi(\frac{2}{3}) - \Phi(-\frac{1}{3}) \approx 0.3779$.

(b) $\mathbb{P}(X > 0) = \mathbb{P}(Z > -1) \approx 0.8413$.

(c) $\mathbb{P}(|X - 3| > 6) = \mathbb{P}(|Z| > 2) = \mathbb{P}(Z > 2) + \mathbb{P}(Z < -2) \approx 0.0456$.

Proposition 2.6. *If X and Y are independent normal random variables with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2) respectively, then $X + Y$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.*

Proof. Observe that,

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{t^2\sigma_1^2/2+t\mu_1}e^{t^2\sigma_2^2/2+t\mu_2} = e^{t^2(\sigma_1^2+\sigma_1^2)/2+t(\mu_1+\mu_2)}.$$

Hence $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. □

Example: Let X and Y be independent standard normal random variables, and let $Z = X + Y$. Calculate the p.d.f of Z.

Solution: $Z \sim N(0, 2)$.

Example: (**Chi-square random variables**) Let Z be a standard normal random variable. Find the distribution of $Y = Z^2$.

Solution: Recall that if Z is a continuous random variable with probability density f_Z then the distribution of $Y = Z^2$ is obtained as follows:

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_Z(\sqrt{y}) + f_Z(-\sqrt{y})].$$

As, $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, we get

$$f_Y(y) = \frac{1}{2\sqrt{y}} \frac{2}{\sqrt{2\pi}} e^{-y/2} = \frac{\frac{1}{2} e^{-y/2} (y/2)^{1/2-1}}{\sqrt{\pi}} = \frac{\frac{1}{2} e^{-y/2} (y/2)^{1/2-1}}{\Gamma(1/2)}.$$

Thus, $Y \sim$ Gamma(1/2,1/2).

Proposition 2.7. *Let $Z_i, i = 1, \ldots, n$ be independent standard normal random variables, then $\sum_{i=1}^n Z_i^2 \sim$ Gamma$(n/2, 1/2)$ or χ^2 with n degrees of freedom.*

Very important in statistical analysis.

2.6 Bivariate normal distribution

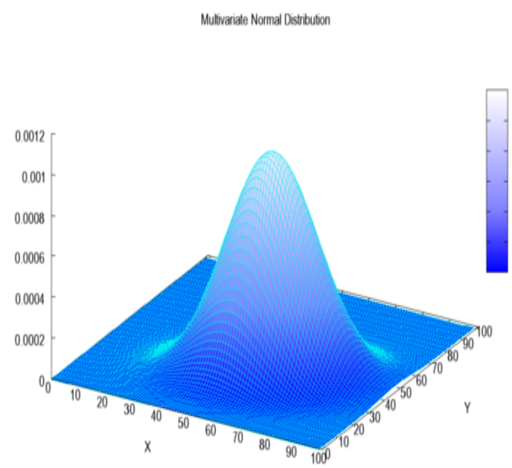
Suppose that Z_1 and Z_2 are independent standard normal r.v's. Let $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ be constants such that $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, and $-1 < \rho < 1$. Define two new random variables X_1 and X_2 as follows:

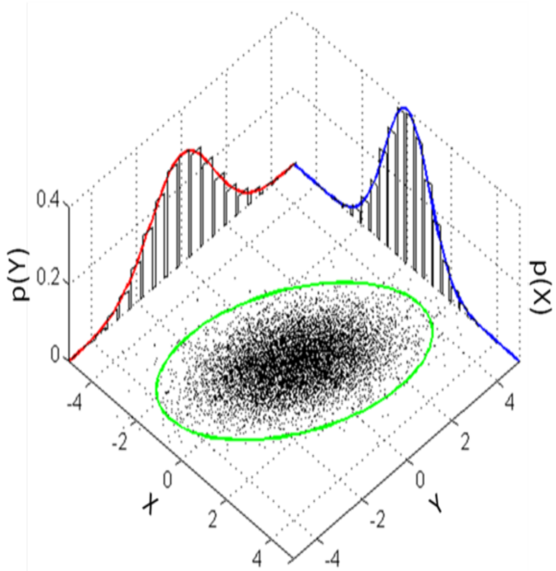
$$\begin{aligned} X_1 &= \sigma_1 Z_1 + \mu_1 \\ X_2 &= \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2) + \mu_2. \end{aligned}$$

Then the joint p.d.f of (X_1, X_2) is

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \, e^{\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]}. \qquad (3)$$

Exercise: Prove this using the multivariate transformation of random variables.





From this representation, we may derive the following key facts: for $i = 1, 2$,

$$\begin{aligned}
\mathbb{E}(X_i) &= \mu_i, \\
\mathrm{Var}(X_i) &= \sigma_i^2, \\
\mathrm{Cor}(X_1, X_2) &= \rho,
\end{aligned}$$

Definition: When the joint p.d.f of (X_1, X_2) is of the form (3) then (X_1, X_2) is said to have a *bivariate normal distribution* with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ.

Contours of this p.d.f are ellipses, centered at (μ_1, μ_2).

Theorem 2.8. *X_1 and X_2 having a joint bivariate normal distribution are independent if and only if $\rho = 0$, i.e., if they are uncorrelated.*

The proof is easy!

Theorem 2.9. *Let (X_1, X_2) have a joint p.d.f of the form in (3). The conditional distribution of X_2 given $X_1 = x_1$ is the normal distribution with mean and variance given by*

$$\mathbb{E}(X_2|X_1 = x_1) = \mu_2 + \rho\sigma_2 \left(\frac{x_1 - \mu_1}{\sigma_1}\right), \qquad Var(X_2|X_1 = x_1) = (1 - \rho^2)\sigma_2^2.$$

Theorem 2.10. *Suppose that (X_1, X_2) have a joint p.d.f of the form in (3). Let $Y = a_1X_1 + a_2X_2 + b$, where $a_1, a_2, b \in \mathbb{R}$ are constants. Then*

$$Y \sim N(a_1\mu_1 + a_2\mu_2 + b, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2).$$

2.7 Multivariate normal

Definition: $\mathbf{X} = (X_1, \ldots, X_d)$ is said to have a joint *multivariate normal distribution* with parameters μ and Σ, written $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right], \qquad \text{for } \mathbf{x} \in \mathbb{R}^d.$$

Here $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is a nonsingular positive definite matrix.

Contours of this p.d.f are ellipses, centered at $\boldsymbol{\mu}$.

$\mathbf{X}$ may be represented as

$$\mathbf{X} = \boldsymbol{R}\mathbf{Z} + \mu,$$

where $\boldsymbol{R}\boldsymbol{R}^\top = \Sigma$ and

$$\mathbf{Z} \sim N_d(\mathbf{0}, I_d)$$

(i.e., $\mathbf{Z}$ is vector with i.i.d standard normal variables).

From this representation, we may derive the following key facts:

$$\begin{aligned} \mathbb{E}(X_i) &= \mu_i, \\ \mathrm{Cov}(X_i, X_j) &= \Sigma_{ij}, \end{aligned}$$

for $i, j \in \{1, 2, \ldots, d\}$.

Note that if $\boldsymbol{A}$ is a $m \times d$ matrix and $\mathbf{b} \in \mathbb{R}^d$, then

$$Y := \boldsymbol{A}\mathbf{X} + \mathbf{b} \sim N_m(\boldsymbol{A}\boldsymbol{\mu} + \mathbf{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top)$$

as $\boldsymbol{A}\mathbf{X} + \mathbf{b} = \boldsymbol{A}\boldsymbol{R}\mathbf{Z} + \boldsymbol{A}\boldsymbol{\mu} + \mathbf{b}$; i.e., *linear transformations preserve Gaussianity.* This also implies that marginalization preserves Gaussianity.