Introduction to Probability Theory for Graduate Economics
Fall 2008

Yiğit Sağlam*

October 10, 2008

CHAPTER 2 - RANDOM VARIABLES AND EXPECTATION[1]

1 Random Variables

A random variable (RV) is a real-valued function over a sample space. Some examples of a RVare the outcome of a fair coin toss (head or tails), the outcome when we roll a dice (an integer between 1 and 6).

In Economics, a RV may represent a wide range of factors: In Job-Search literature, the value of a wage offer is a RV. In Macroeconomics, production of a commodity may include a stochastic component, such as total factor productivity, and this stochastic part is a RV. In Auctions literature, bidders have a private or a common value which also is a RV defined over a specified domain.

Random variables are commonly represented by capital letters, whereas lowercase letters denote a realization of a RV. For example, X represents the outcome of a coin toss, where x equals the particular outcome *heads*. Random variables are of two types: discrete, and continuous. It is noteworthy that a random may be a mixture of both discrete and continuous RVs. The feature *Discrete* or *Continuous* describes the nature of the domain (or the sample space) of the RV. For instance, the outcome of a coin toss is a discrete RV, while the value of wage offer is a continuous RV.

Since a RV has multiple outcomes, one needs a probability model for observing those outcomes. So, the ***probability distribution*** of a RV provides information about what values the RV can take and how to assign probabilities to those values.

The probability mass function (pmf) of a *discrete* RV X lists the possible values x_i for the RV and their respective probabilities p_i. The only restrictions on the probabilities are the following:

- Every probability must be positive and less than 1:

$$0 \leq p_i \leq 1; \forall\ i = 1, 2, \ldots, n.$$

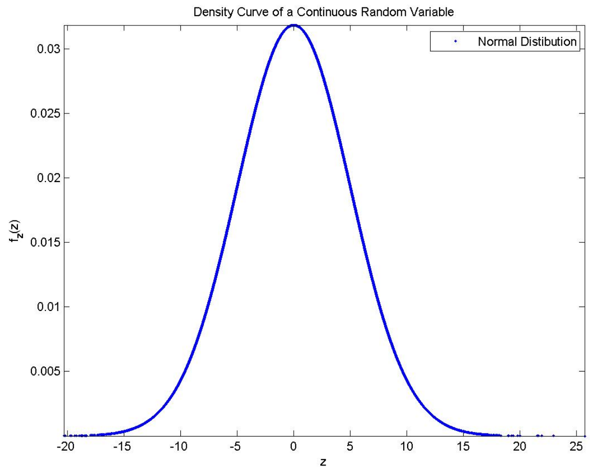
- The sum of the probabilities must equal 1:

$$\sum_{i=1}^{n} p_i = 1.$$

*PhD Candidate in Economics, Department of Economics, University of Iowa, W210 Pappajohn Business Building, Iowa City, IA 52242, yigit-saglam@uiowa.edu, Phone: 1(319)3350829, Fax: 1(319)3351956

[1] I would like to thank Brent Hickman for his helpful comments.

Figure 1 - Density Curve of a Continuous Random Variable



The probability distribution of a *continuous* RV X is described by a density curve. The probability of any event is the area under *the density curve* and above the values of X that make up the event. Suppose a RV X may take all values over an interval of real numbers. Then the probability that X is in the set of outcomes A, $f_X(A)$, is defined to be the area above A and under a curve. The curve, which represents a function $f_X(X)$, must satisfy the following:

- The curve has no negative values:
$$p(x) > 0; \forall\ x.$$
- The total area under the curve is equal to 1:
$$\int_A f_X(x)\ dx = 1.$$

2 Discrete Random Variables

2.1 Bernoulli Distribution

A Bernoulli Experiment is a random experiment the outcome of which is one of the two mutually exclusive events such as failure and success, or head and tail, etc.. Denote the probability for these outcomes; ie, success and failure, are θ and $(1-\theta)$. The probabilities remain the same over trials.

Let X be a Bernoulli RV. Furthermore, let X equal 1 if the trial is a success, and X equal 0 if the trial is a failure. The probability mass function (pmf), the mean, and the variance of X are:

$$f(x;\theta) = \theta^x\ (1-\theta)^{1-x}; \forall x = 0, 1$$
$$\mu_x = \mathcal{E}(x) = \sum_{x=0}^{1} x\ \theta^x\ (1-\theta)^{1-x} = \theta,$$
$$\sigma_x^2 = Var(x) = \sum_{x=0}^{1} (x-\theta)^2\ \theta^x\ (1-\theta)^{1-x} = \theta(1-\theta).$$

Applications:

- In the Asset Pricing literature, in the case of discontinuous interest rates, the jump in the interest rate is a Bernoulli RV which either occurs in a period with a certain probability or not; see Das (2002).

2.2 Binomial Distribution

Binomial distribution is general case of Bernoulli distribution when there is a fixed number of trials. Let n denote the number of trials. The trials are independent, and the probabilities remain the same over trials.

Let X be a Binomial RV that shows the number of successes in n trials. Since the probability of success is θ and the failure is $(1-\theta)$ for each trial, the probability mass function of X is the sum of probabilities of $\binom{n}{x}$ mutually exclusive events.

The probability mass function (pmf), the mean, and the variance of X are:

$$f(x;\theta,n) = \binom{n}{x}\ \theta^x\ (1-\theta)^{n-x}; \forall x = 0, 1, \ldots, n$$

$$\mu_x = \mathcal{E}(x) = \sum_{x=0}^{n} \binom{n}{x}\ x\ \theta^x\ (1-\theta)^{n-x} = n\theta,$$

$$\sigma_x^2 = Var(x) = \sum_{x=0}^{n} \binom{n}{x}\ (x-n\theta)^2\ \theta^x\ (1-\theta)^{n-x} = n\theta(1-\theta).$$

It is noteworthy that one can derive the Geometric and Hypergeometric distributions using the Bernoulli distribution.

Applications:

- In the Labor Supply literature, the number of days a person is absent from work is a Binomial RV assuming that the days are not serially correlated; see Johansson and Palme (1996). In their paper, under some assumptions, Johansson and Palme maximized the Binomial log likelihood function to estimate the parameters.

Exercise 1.1 Suppose that an airplane engine will fail, when in flight, with probability $(1-\theta)$ independently from engine to engine; suppose that the airplane will make a successful flight if at least 50 percent of its engines remain operative. For what values of θ is a four-engine plane preferable to a two-engine plane?

Answer 1.1 As each engine is assumed to fail or function independently of what happens with the other engines, it follows that the number of engines remaining operative is a binomial RV. Hence, the probability that a four-engine plane makes a successful flight is:

$$\begin{aligned} Pr(X \geq 2) &= \binom{4}{2}\ \theta^2\ (1-\theta)^{4-2} + \binom{4}{3}\ \theta^3\ (1-\theta)^{4-3} + \binom{4}{4}\ \theta^4\ (1-\theta)^{4-4} \\ &= 6\theta^2\ (1-\theta)^2 + 4\theta^3\ (1-\theta) + \theta^4 \end{aligned}$$

whereas the corresponding probability for a two-engine plane is:

$$Pr(X \geq 1) = \binom{2}{1}\ \theta^1\ (1-\theta)^{2-1} + \binom{2}{2}\ \theta^2\ (1-\theta)^{2-2} = 2\theta^1\ (1-\theta)^1 + \theta^2$$

Hence, the four-engine plane is safer if:

$$\begin{aligned} &\Rightarrow 6\theta^2\ (1-\theta)^2 + 4\theta^3\ (1-\theta) + \theta^4 \geq 2\theta^1\ (1-\theta)^1 + \theta^2 \\ &\Rightarrow 6\theta^1\ (1-\theta)^2 + 4\theta^2\ (1-\theta) + \theta^3 \geq 2-\theta \\ &\Rightarrow \theta \geq \frac{2}{3}\ \blacksquare \end{aligned}$$

2.3 Multinomial Distribution

The multinomial distribution is a generalization of the binomial distribution. The binomial distribution is the probability distribution of the number of *successes* in n independent Bernoulli trials, with the same probability of *success* on each trial. Instead of each trial resulting in *success* or *failure*, we assume that each trial results in one of some fixed finite number k of possible outcomes, with probabilities $p_1, \ldots, p_k$, and there are n independent trials. We can use a RV X_i to indicate the number of times outcome number i was observed over the n trials.

The probability mass function (pmf), the mean, and the variance of $X_1, \ldots, X_k$ are:

$$f(x_1, \ldots, x_k; p_1, \ldots, p_k, n) = \begin{cases} \binom{n}{x_1, \ldots, x_k} \prod_{i=1}^{k} p_i^{x_i} & \text{if} \sum_{i=1}^{k} x_i = n, \\ 0 & \text{otherwise.} \end{cases}$$

$$\mu_{x_i} = \mathcal{E}(x_i) = np_i; \forall\ i = 1, \ldots, k,$$

$$\sigma_{x_i} = Var(x_i) = np_i(1 - p_i); \forall\ i = 1, \ldots, k.$$

It is noteworthy that multinomial distribution is a joint distribution, which will be covered more thoroughly in Chapter 4. However, it is introduced here merely as another discrete probability distribution.

2.4 Poisson Distribution

Poisson distribution is particularly useful for modeling time instants at which events occur. Assume that we are going to observe the interested event for a period of time T. The time instant at which we start to observe the events will be labeled "0", the origin of time scale. The number of events in this time interval $(0, T)$ is a RV X.

The probability mass function (pmf), the mean, and the variance of X are:

$$f(x; \lambda) = \frac{\lambda^x\ e^{-\lambda}}{x!}; \forall x = 0, 1, \ldots$$

$$\mu_x = \mathcal{E}(x) = \sum_{x=0}^{\infty} x\ \frac{\lambda^x\ e^{-\lambda}}{x!} = \lambda,$$

$$\sigma_x^2 = Var(x) = \sum_{x=0}^{\infty} (x - \lambda)^2\ \frac{\lambda^x\ e^{-\lambda}}{x!} = \lambda.$$

Approximation to Binomial Distribution:

An important property of the Poisson rv is that it may be used to approximate a binomial RV when the number of trials n is large, and the probability of success θ is small. To see this, let $\lambda = n\theta$, and consider the following equations:

$$\begin{aligned} Pr(X = i) &= \frac{n!}{(n-i)!\ i!} \theta^i (1-\theta)^{n-i} \\ &= \frac{n!}{(n-i)!\ i!} \left(\frac{\lambda}{n}\right)^i (1 - \frac{\lambda}{n})^{n-i} \\ &= \frac{n(n-1)\ldots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}; \text{for } i = 0, 1, \ldots, n. \end{aligned}$$

For n large, θ small:

$$(1 - \lambda/n)^n \approx \exp(-\lambda);\ \frac{n(n-1)\ldots(n-i+1)}{n^i} \approx 1;\ (1 - \lambda/n)^i \approx 1$$

Finally, we can approximate the probability in the following way:

$$Pr(X = i) \approx \exp(-\lambda) \frac{\lambda^i}{i!}.$$

Applications:

- There is a certain relationship between Exponential and Poisson distributions. To wit, exponential distribution can be derived from Poisson distribution. We will explore this relationship in the section for the Exponential distribution. Exponential and Poisson distribution are both useful in modeling stochastic processes with certain features. In particular, exponentially distributed RVs are "memoryless processes". As an example from the Industrial Organization literature, the distribution of increments for an innovation can be a Poisson distribution, if it a "memoryless process"; see Hopenhayn and Squintani (2004). We will see more about the stochastic processes in Chapter 5.
- In the Labor literature, the distribution of job offers in a specific time length can be a Poisson distribution; see Van Den Berg (1990).

Exercise 1.2 Suppose that the number of typological errors on a single page of a book has a Poisson distribution with parameter $\lambda = 1$. Calculate the probability that there is at least one errors on a page.

Answer 1.2 To calculate the probability that a page has at least one errors, we first calculate the probability that there is no error on the page:

$$\begin{aligned} Pr(X \geq 1) = 1 - Pr(X = 0) = 1 - \left[\exp(-\lambda) \frac{\lambda^0}{0!} \right] \\ = 1 - \exp(-1) \approx 0.6333 \; \blacksquare \end{aligned}$$

Exercise 1.3 If the number of wage offers that a person gets each day is a Poisson RV with parameter $\lambda = 3$, what is the probability that no wages are offered today?

Answer 1.3 To calculate the probability that a person has no wage offers:

$$\begin{aligned} Pr(X = 0) = \left[\exp(-\lambda) \frac{\lambda^0}{0!} \right] \\ = \exp(-3) \approx 0.05 \; \blacksquare \end{aligned}$$

3 Continuous Random Variables

3.1 Uniform Distribution

The uniform distribution defines equal probability over a given range for a continuous distribution. For this reason, it is important as a reference distribution.

The probability distribution function, mean, and variance of X are:

$$f(x) = \frac{1}{B-A}; \forall\ x \in [A, B],$$
$$\mu_x = \mathcal{E}(x) = \int_A^B x \left(\frac{1}{B-A}\right) dx = \frac{A+B}{2},$$
$$\sigma_x^2 = Var(x) = \int_A^B \left(x - \frac{A+B}{2}\right)^2 \left(\frac{1}{B-A}\right) dx = \frac{(B-A)^2}{12}.$$

Applications:

- One of the most important applications of the uniform distribution is in the generation of random numbers. That is, almost all random number generators generate random numbers on the [0,1] interval. For other distributions, some transformation is applied to the uniform random numbers. This is called Inverse Transform Sampling Method. The procedure is as follows: First, one generates random draws y from the uniform distribution defined over [0,1]. Then, if the cumulative distribution function is known and continuous over the domain, then one can find the value x for which the cumulative probability equals y. This method is very useful in theoretical work. However, this method may not work efficiently for some distributions such as the normal distribution. More details about this method will be covered in Chapter 3.

3.2 Exponential Distribution

We developed the distribution of the number of occurrences in the interval $(0, T)$, which is a Poisson distribution with parameter λ. Now, let T be the time at which the first event occurs. Then, the RV T is continuous. Consider the event $T > t$, that the time of the first event is greater than t. The probability that there is zero event until time t becomes:

$$Pr\,(T > t) = Pr\,(x = 0; \lambda) = \frac{\lambda^0\ e^{-\lambda}}{0!} = e^{-\lambda} = e^{-ct}$$

where λ equals ct, proportional to t. Then, the cumulative distribution of T is:

$$F_T(T) = Pr\,(T < t) = 1 - e^{-ct}$$

The probability distribution function, mean, and variance of T are:

$$f(t;\theta) = \begin{cases} \frac{1}{\theta}\ e^{\frac{-t}{\theta}}; & \forall x > 0 \\ 0 & \text{otherwise} \end{cases}$$
$$\mu_x = \mathcal{E}(x) = \int_0^\infty t\ \frac{1}{\theta}\ e^{\frac{-t}{\theta}} dt = \theta,$$
$$\sigma_x^2 = Var(x) = \int_0^\infty (t-\theta)^2\ \frac{1}{\theta}\ e^{\frac{-t}{\theta}} dt = \theta^2.$$

Applications:

- A very important feature of the Exponential distribution is that it is a memoryless process. To wit, let X be an Exponentially distributed RV, and consider the probability that $X \geq t$.

$$Pr(X \geq t) = Pr\left(X \geq t_o + t | X \geq t_o\right); \ \forall t_o \geq 0, \text{and for } t > 0.$$

  In words, the probability that the first occurrence happens at a time $X \geq t$ is equivalent to the probability that the first occurrence happens at time $X \geq t_o + t$, given that it has not yet occurred until time t_o. Whenever it is appropriate, the memoryless property of Exponential distribution is useful in economics, as one does not have to keep track of the whole history to compute the probability distribution of a variable in the current state.

3.3 Gamma Distribution

Suppose that we wish to examine the continuous RV X measuring the required length for r events to occur. Suppose that we are observing a Poisson process starting at time 0 and let T_r be the time of an occurrence of the r^{th} event. Also, let t any fixed positive number and consider the event $T_r > t$; ie, the time of the r^{th} event is greater than t. This event $\{T_r > t\}$ is equivalent to $\{X \leq r - 1\}$, where X is the number of events that occur in $[0, t]$. Thus, X is a Poisson RV with parameter $\lambda = ct$:

$$\begin{aligned}
Pr\left(T_r > t\right) &= Pr\left(X \leq r - 1\right) = \sum k = 0^{r-1} \exp(-\lambda) \frac{\lambda^k}{k!} \\
F_{T_r}(t) &= 1 - Pr\left(T_r \geq t\right) = 1 - \sum k = 0^{r-1} \exp(-ct) \frac{(ct)^k}{k!} \\
f_{T_r}(t) &= \frac{c^r t^{r-1}}{(r-1)!} \ \exp(-ct), \ \forall t > 0
\end{aligned}$$

where $f_{T_r}(t)$ is the pdf of X, and it is still discrete due to r. This is a special case of Gamma Probability Law and is called Erlang Law. However, since our aim is to make this expression valid for any positive real number r, we will employ the Gamma Function $\Gamma(r)$. To derive the pdf of a Gamma RV, let $y = x/\beta$ and consider the following:

$$\begin{aligned}
\Gamma(r) &= \int_0^\infty y^{r-1} \exp(-y) dy; \ \text{for any } r > 0. \\
f(x) &= \frac{x^{\alpha-1} \exp(-x/\beta)}{\Gamma(\alpha)\beta^\alpha}
\end{aligned}$$

where $\{r, t, c\}$ equal $\{\alpha, x, \beta^-1\}$. Finally, α and β are usually known as shape and scale parameters, respectively. One can also see the Gamma distribution in the following way: The sum of independently and exponentially distributed RVs is a Gamma distributed RV.

The probability distribution function, mean, and variance of X are:

$$\begin{aligned}
f(x, \alpha, \beta) &= \frac{x^{\alpha-1} \exp(-x/\beta)}{\Gamma(\alpha)\beta^\alpha}; \text{for } r > 0, \\
\mu_x = \mathcal{E}(x) &= \int_0^\infty x \ f(x, \alpha, \beta) \ dx = \alpha\beta, \\
\sigma_x^2 = Var(x) &= \int_0^\infty (x - \alpha\beta)^2 \ f(x, \alpha, \beta) \ dx = \alpha\beta^2.
\end{aligned}$$

Applications:

- Chi-square distribution is a special case of Gamma distribution:

$$X \sim Gamma\left(\alpha = 1/2, 1, 3/2, 2, 5/2, \ldots, \beta = 2\right)$$

- Exponential distribution, which we will cover next, is a special case of Gamma distribution:

$$X \sim Gamma\left(\alpha = 1, \beta\right)$$

3.4 Normal Distribution

The normal distribution, also called the Gaussian distribution, is an important family of continuous probability distributions. For both theoretical and practical reasons, the normal distribution is probably the most important distribution in statistics. For example, many classical statistical tests are based on the assumption that the data follow a normal distribution. In modeling applications, such as linear and non-linear regression, the error term is often assumed to follow a normal distribution with fixed location and scale. Also, the normal distribution is used for inference; ie, to find significance levels in many hypothesis tests and confidence intervals.

The probability distribution function, mean, and variance of X are:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \forall\ x \in [A, B],$$

$$\mu_x = \mathcal{E}(x) = \int_{-\infty}^{\infty} x\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu,$$

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

Applications:

- The normal distribution is widely used. Part of the appeal is that it is well behaved and mathematically tractable. However, the central limit theorem provides a theoretical basis for why it has wide applicability. The central limit theorem basically states that as the sample size (N) becomes large, the following occur:
  - * The sampling distribution of the mean becomes approximately normal regardless of the distribution of the original variable.
  - * The sampling distribution of the mean is centered at the population mean, μ, of the original variable. In addition, the standard deviation of the sampling distribution of the mean approaches $\sigma/\sqrt{n}$.
- As mentioned above, normal distribution is widely used in econometrics in linear and non-linear regression, and in inference.

3.5 Log-Normal Distribution

The log-normal distribution is the single-tailed probability distribution of any random variableRV whose logarithm is normally distributed. If X is a RV with a normal distribution, then $Y = \exp(X)$ has a log-normal distribution; likewise, if Y is log-normally distributed, then $\log(Y)$ is normally distributed.

The probability distribution function, mean, and variance of X are:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}; \forall\ x \in [A, B],$$

$$\mu_x = \mathcal{E}(x) = \int_{-\infty}^{\infty} x\ \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} dx = e^{\mu+\sigma^2/2},$$

$$\sigma_x^2 = Var(x) = \int_{-\infty}^{\infty} \left(x - e^{\mu+\sigma^2/2}\right)^2\ \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} dx = \left(e^{\sigma^2} - 1\right)\ e^{2\mu+\sigma^2}.$$

Applications:

- In Macroeconomics, the productivity shock to the production function of a commodity is taken as a log-normal RV. This assumption may be convenient as the log-normal distribution is defined over the positive real numbers, and the distribution function is tractable.

4 Expectation, Moments, and Moment Generating Functions

4.1 Expectation

Definition ***Expectation*** Let X be a RV.

- If X is a continuous RV with a pdf $f(x)$ and $\int_{-\infty}^{\infty} |x|\, f(x)\, dx < \infty$, then the **expectation** of X is
$$\mathcal{E}(X) = \int_{-\infty}^{\infty} x\ f(x)\ dx.$$
- If X is a discrete RV with a pmf $f(x)$ and $\sum_x |x|\ f(x) < \infty$, then the **expectation** of X is
$$\mathcal{E}(X) = \sum_x x\ f(x) \blacksquare$$

Expectation of X is also known as the **expected value**, or the **mean** of X.

Theorem Let X be a RV, and $g(.)$ be a function of X. Then:

- If X is a continuous RV with a pdf $f(x)$ and $\int_{-\infty}^{\infty} |g(x)|\ f(x)\ dx < \infty$, then
$$\mathcal{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\ f(x)\ dx.$$
- If X is a discrete RV with a pmf $f(x)$ and $\sum_x |g(x)|\ f(x) < \infty$, then
$$\mathcal{E}(g(X)) = \sum_x g(x)\ f(x) \blacksquare$$

Theorem ***Expectation is a linear operator.*** Let X be a RV, and $g_1(.)$ and $g_2(.)$ be functions of X. Suppose that the expectations of $g_1(X)$ and $g_2(X)$ exist. Then, for any constants k_1 and k_2, the expectation of $k_1\ g_1(X) + k_2\ g_2(X)$ exists and is given by::
$$\mathcal{E}\left[k_1\ g_1(X) + k_2\ g_2(X)\right] = k_1\ \mathcal{E}\left[g_1(X)\right] + k_2\ \mathcal{E}\left[g_2(X)\right].$$

Proposition ***Law of Total Expectation (or Law of Iterated Expectations)*** Let X and Y be two RVs. Define the conditional expectation
$$\mathcal{E}(X|Y)(y) = \mathcal{E}(X|Y = y).$$

Then the expectation of X satisfies
$$\mathcal{E}(X) = \mathcal{E}\left[\mathcal{E}(X|Y)\right].$$

Proof: Discrete Case:

$$\begin{aligned}
\mathcal{E}\left[\mathcal{E}(X|Y)\right] &= \sum_y \mathcal{E}(X|Y=y)\ Pr(Y=y) \\
&= \sum_y \left(\sum_x x\ Pr(X=x|Y=y)\right)\ Pr(Y=y) \\
&= \sum_y \sum_x x\ Pr(X=x|Y=y)\ Pr(Y=y) \\
&= \sum_x \sum_y x\ Pr(Y=y|X=x)\ Pr(X=x) \\
&= \sum_x x\ Pr(X=x)\ \left(\sum_y Pr(Y=y|X=x)\right) \\
&= \sum_x x\ Pr(X=x) \\
&= \mathcal{E}(X)\blacksquare
\end{aligned}$$

4.2 Some Special Expectations

Definition ***Mean*** Let X be a RV whose expectation exists. The mean value of X is the defined to be

$$\mu = \mathcal{E}(X).$$

Definition ***Variance*** Let X be a RV with finite mean μ and such that $\mathcal{E}\left[(x-\mu)^2\right]$ exists. Then, the variance of X is defined to be $\sigma^2 = \mathcal{E}\left[(x-\mu)^2\right] = Var(X)$. Moreover,

$$Var(X) = \sigma^2 = \mathcal{E}(X^2) - [\mathcal{E}(X)]^2 = \mathcal{E}(X^2) - \mu^2.$$

Definition ***Covariance*** Covariance is a measure of how much two variables change together. In particular, the variance is a special case of the covariance when the two variables are identical. Let X and Y be two RVs whose expectations exist. The covariance of X and Y is the defined to be

$$Cov(X,Y) = \mathcal{E}(XY) - \mathcal{E}(X)\ \mathcal{E}(Y) = \mathcal{E}(XY) - \mu_X\ \mu_Y.$$

It is noteworthy that if two RVs are independent, then the covariance equals 0. However, the converse is not true in general.

Definition ***Moment Generating Functions*** Let X be a RV such that for some $h > 0$, the expectation of e^{tX} exists for $-h < t < h$. The moment generating function of X (henceforth: **mgf**) is defined to be the function

$$M(t) = \mathcal{E}(e^{tX});\ \text{for } t \in [-h, h].$$

Theorem: ***While a distribution has a unique mgf, mgf's uniquely identify distributions.*** Let X and Y be two RVs with mgf's M_X and M_Y, respectively, existing in open interval about 0. Then

$$F_X(z) = F_Y(z);\ \forall z \in \mathcal{R} \Longleftrightarrow M_X(t) = M_Y(t);\ \forall t \in [-h, h] \text{ for some } h > 0 \blacksquare$$

A nice feature of the mgf's is that one can derive the mean and variance using the mgf's. To wit:

$$\begin{aligned}
M(t) &= \mathcal{E}(X), \\
M'(t) &= \frac{d}{dt} M(t) = \mathcal{E}\left(\frac{d}{dt} e^{tX}\right) = \mathcal{E}\left(X\ e^{tX}\right), \\
\Rightarrow M'(0) &= \mathcal{E}(X) = \mu, \\
\Rightarrow M''(0) &= \mathcal{E}(X^2) = \sigma^2 + \mu^2, \\
&\vdots \\
\Rightarrow M^{(m)}(0) &= \mathcal{E}(X^m) = \begin{cases} \int_{-\infty}^{\infty} x^m\ f(x)\ dx & \text{if } X \text{ is a continuous RV,} \\ \sum_x x^m\ f(x) & \text{if } X \text{ is a discrete RV} \end{cases}
\end{aligned}$$

where $M^m(0)$ is the m^{th} derivative of the mgf of X evaluated at $t = 0$, which is also known as the m^{th} *raw moment* around the origin.

Definition ***Central Moments*** The m^{th} central moment of the probability distribution of a random variableRV X is the moment around the mean and is denoted in the following way:

$$\mu_m = \mathcal{E}\left[(X - \mu)^m\right].$$

1. $\mu_1 = \mathcal{E}\left[(X - \mu)^1\right] = 0.$

2. **Variance** $= Var(X) = \sigma^2 = \mu_2 = \mathcal{E}\left[(X - \mu)^2\right].$

3. **Skewness** $= \mu_3 = \mathcal{E}\left[(X - \mu)^3\right].$

4. **Kurtosis** $= \mu_4 = \mathcal{E}\left[(X - \mu)^4\right].$

4.3 Some Examples for Moment Generating Functions

Exercise 3.1 ***MGF of Binomial Distribution with Parameters*** n ***and*** θ

$$\begin{aligned}
M(t) = \mathcal{E}(e^{tX}) &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} \theta^k \ (1-\theta)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} \left(\theta e^t\right)^k \ (1-\theta)^{n-k} \\
&= \left(\theta e^t + 1 - \theta\right)^n . \\
\mathcal{E}(X) = \mu = M'(t=0) &= n\left(\theta e^0 + 1 - \theta\right)^{n-1} \ \theta \ e^0 = n \ \theta. \\
\mathcal{E}(X^2) = M''(t=0) &= n \ (n-1) \ \theta^2 + n \ \theta. \\
Var(X) = \sigma^2 &= n \ \theta \ (1-\theta).
\end{aligned}$$

Exercise 3.2 ***MGF of Poisson Distribution with Parameter*** λ

$$\begin{aligned}
M(t) = \mathcal{E}(e^{tX}) &= \sum_{k=0}^{\infty} \frac{e^{tk} \ e^{-\lambda} \ \lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^t\right)^k}{k!} \\
&= e^{-\lambda} \ e^{\lambda e^t} = e^{\lambda\left(e^t - 1\right)}. \\
\mathcal{E}(X) = \mu = M'(t=0) &= \lambda \ e^0 \ e^{\lambda\left(e^0-1\right)} = \lambda. \\
\mathcal{E}(X^2) = M''(t=0) &= \left[\left(\lambda e^0\right)^2 + \lambda e^0\right] \ e^{\lambda\left(e^0-1\right)} = \lambda^2 + \lambda. \\
Var(X) = \sigma^2 &= \lambda.
\end{aligned}$$

Exercise 3.3 ***MGF of Exponential Distribution with Parameter*** θ

$$\begin{aligned}
M(t) = \mathcal{E}(e^{tX}) &= \int_0^{\infty} e^{tx} \ \frac{e^{-x/\theta}}{\theta} \ dx \\
&= \frac{1}{\theta} \int_0^{\infty} e^{-(1/\theta - t)x} \ dx \\
&= \frac{1}{1-\theta \ t}; \ \text{for } t < \theta^{-1}. \\
\mathcal{E}(X) = \mu = M'(t=0) &= \frac{\theta}{(1-\theta \ 0)^2} = \theta. \\
\mathcal{E}(X^2) = M''(t=0) &= \frac{2\theta^2}{(1-\theta \ 0)^3} = 2\theta^2. \\
Var(X) = \sigma^2 &= \theta^2.
\end{aligned}$$

Exercise 3.4 ***MGF of Normal Distribution with Parameters*** μ ***and*** σ^2

$$
\begin{aligned}
M(t) = \mathcal{E}(e^{tX}) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{tx}\ e^{\frac{-(x-\mu)^2}{2\sigma^2}}\ dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[\frac{-\left(x^2 - 2\mu x\mu^2 - 2\sigma^2 tx\right)}{2\sigma^2}\right]\ dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[\frac{-\left(x-\left(\mu+\sigma^2 t\right)\right)^2 + \sigma^4 t^2 + 2\mu\sigma^2 t}{2\sigma^2}\right]\ dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{\sigma^4 t^2 + 2\mu\sigma^2 t}{2\sigma^2}\right] \int_{-\infty}^{\infty} \exp\left[\frac{-\left(x-\left(\mu+\sigma^2 t\right)\right)^2}{2\sigma^2}\right]\ dx \\
&= \exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right] \left\{\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[\frac{-\left(x-\left(\mu+\sigma^2 t\right)\right)^2}{2\sigma^2}\right]\ dx\right\} \\
&= \exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right] \{1\} \\
&= \exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right].
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{E}(X) = \mu = M'(t=0) &= \left(\mu + \sigma^2 0\right)\ \exp\left[\frac{\sigma^2 0^2}{2} + \mu 0\right] = \mu. \\
\mathcal{E}(X^2) = M''(t=0) &= \left[\left(\mu + \sigma^2 0\right)^2 + \sigma^2\right]\ \exp\left[\frac{\sigma^2 0^2}{2} + \mu 0\right] = \mu^2 + \sigma^2. \\
Var(X) &= \sigma^2 = \sigma^2.
\end{aligned}
$$

Exercise 3.5 ***MGF of LogNormal Distribution does <u>not</u> exist.*** Even though the LogNormal distribution has finite moments of all orders, the moment generating function is infinite at any positive number. This property is one of the reasons for the fame of the LogNormal distribution.

$$
M(t) = \mathcal{E}(e^{tX}) \to \infty;\ \text{for any } t > 0.
$$

5 References

Das, Sanjiv R. "The Surprise Element: Jumps in Interest Rates." *Journal of Econometrics*, 106 (2002), 27–65.

Johansson, P. and M. Palme. "Do Economic Incentives Affect Work Absence? Empirical Evidence Using Swedish Micro Data." *Journal of Public Economics*, 59 (1996), 195–218.

Squintani, F. and H. Hopenhayn. "Preemption Games with Private Information." 2005 Meeting Papers 80, Society for Economic Dynamics.

Van Der Berg, Gerard J. "NonStationarity in Job Search Theory." *Review of Economic Studies*, 57 (1990), 255-277.

Ross, Sheldon M. *Introduction to Probability Models.* Fifth Edition. San Diego: Academic Press, 1993.