LECTURE 4:
COMMONLY USED DISTRIBUTIONS

VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

Section 4.1: The Bernoulli Distribution

We use the Bernoulli distribution when we have an experiment which can result in one of two outcomes. One outcome is labeled "success," and the other outcome is labeled "failure."

The probability of a success is denoted by p. The probability of a failure is then $1 - p$.

Such a trial is called a **Bernoulli trial** with success probability p.

Examples 1 and 2

1. The simplest Bernoulli trial is the toss of a coin. The two outcomes are heads and tails. If we define heads to be the success outcome, then p is the probability that the coin comes up heads. For a fair coin, $p = 0.5$.
2. Another Bernoulli trial is a selection of a component from a population of components, some of which are defective. If we define "success" to be a defective component, then p is the proportion of defective components in the population.

$X \sim \text{Bernoulli}(p)$

For any Bernoulli trial, we define a random variable X as follows:

If the experiment results in a success, then $X = 1$. Otherwise, $X = 0$. It follows that X is a discrete random variable, with probability mass function $p(x)$ defined by

$$p(0) = P(X = 0) = 1 - p$$

$$p(1) = P(X = 1) = p$$

$$p(x) = 0 \text{ for any value of } x \text{ other than 0 or 1}$$

Mean and Variance

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

- $\mu_X = 0(1-p) + 1(p) = p$

- $\sigma_X^2 = (0-p)^2(1-p) + (1-p)^2(p) = p(1-p)$

Example 3

Ten percent of components manufactured by a certain process are defective. A component is chosen at random. Let $X = 1$ if the component is defective, and $X = 0$ otherwise.

1. What is the distribution of X?

2. Find the mean and variance of X.

Solution

The success probability is $p = P(X = 1) = 0.1$. Therefore $X \sim \text{Bernoulli}(0.1)$.

Solution

Since $X \sim \text{Bernoulli}(0.1)$, the success probability p is equal to 0.1. Using Equations (4.1) and (4.2), $\mu_X = 0.1$ and $\sigma_X^2 = 0.1(1 - 0.1) = 0.09$.

Section 4.2: The Binomial Distribution

If a total of n Bernoulli trials are conducted, and

- The trials are independent.
- Each trial has the same success probability p
- X is the number of successes in the n trials

then X has the **binomial distribution** with parameters n and p, denoted $X \sim \text{Bin}(n, p)$.

Example 4

A fair coin is tossed 10 times. Let X be the number of heads that appear. What is the distribution of X?

Solution

There are 10 independent Bernoulli trials, each with success probability $p = 0.5$. The random variable X is equal to the number of successes in the 10 trials. Therefore $X \sim \text{Bin}(10, 0.5)$.

Another Use of the Binomial

Assume that a finite population contains items of two types, successes and failures, and that a simple random sample is drawn from the population. Then if the sample size is no more than 5% of the population, the binomial distribution may be used to model the number of successes.

Example 5

A lot contains several thousand components, 10% of which are defective. Seven components are sampled from the lot. Let X represent the number of defective components in the sample. What is the distribution of X?

Solution

Since the sample size is small compared to the population (i.e., less than 5%), the number of successes in the sample approximately follows a binomial distribution. Therefore we model X with the Bin(7, 0.1) distribution.

Binomial R.V.: pmf, mean, and variance

- If $X \sim \text{Bin}(n, p)$, the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, x = 0,1,\ldots,n \\ 0, \quad \text{otherwise} \end{cases}$$

- Mean: $\mu_X = np$

- Variance:

$$\sigma_X^2 = np(1-p)$$

Example 6

A large industrial firm allows a discount on any invoice that is paid within 30 days. Of all invoices, 10% receive the discount. In a company audit, 12 invoices are sampled at random. What is the probability that fewer than 4 of the 12 sampled invoices receive the discount?

Solution

Let X represent the number of invoices in the sample that receive discounts. Then $X \sim \text{Bin}(12, 0.1)$. The probability that fewer than four invoices receive discounts is $P(X \leq 3)$. We consult Table A.1 with $n = 12$, $p = 0.1$, and $x = 3$. We find that $P(X \leq 3) = 0.974$.

More on the Binomial

- Assume n independent Bernoulli trials are conducted.
- Each trial has probability of success p.
- Let $Y_1, \ldots, Y_n$ be defined as follows: $Y_i = 1$ if the i^{th} trial results in success, and $Y_i = 0$ otherwise. (Each of the Y_i has the Bernoulli(p) distribution.)
- Now, let X represent the number of successes among the n trials. So, $X = Y_1 + \ldots + Y_n$.

> This shows that a binomial random variable can be expressed as a sum of Bernoulli random variables.

Estimate of *p*

If $X \sim \text{Bin}(n, p)$, then the sample proportion $\hat{p} = X / n$ is used to estimate the success probability p.

Note:

- Bias is the difference $\mu_{\hat{p}} - p$.
- $\hat{p}$ is unbiased.
- The uncertainty in $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

- In practice, when computing σ, we substitute $\hat{p}$ for p, since p is unknown.

Example 7

In a sample of 100 newly manufactured automobile tires, 7 are found to have minor flaws on the tread. If four newly manufactured tires are selected at random and installed on a car, estimate the probability that none of the four tires have a flaw, and find the uncertainty in this estimate.

Solution

Let p represent the probability that a tire has no flaw. The probability that all four tires have no flaw is p^4. We use propagation of error (Section 3.3) to estimate the uncertainty in p^4. We begin by computing the sample proportion $\widehat{p}$ and finding its uncertainty. The sample proportion is $\widehat{p} = 93/100 = 0.93$. The uncertainty in $\widehat{p}$ is given by $\sigma_{\widehat{p}} = \sqrt{p(1-p)/n}$. We substitute $n = 100$ and $\widehat{p} = 0.93$ for p to obtain

We estimate p^4 with $\widehat{p}^4 = 0.93^4 = 0.7481$. We use Equation (3.10) to compute the uncertainty in $\widehat{p}^4$:

$$
\begin{aligned}
\sigma_{\widehat{p}^4} &\approx \left| \frac{d}{d\widehat{p}} \widehat{p}^4 \right| \sigma_{\widehat{p}} \\
&= 4\widehat{p}^3 \sigma_{\widehat{p}} \\
&= 4(0.93)^3(0.0255) \\
&= 0.082
\end{aligned}
$$

Section 4.3: The Poisson Distribution

- One way to think of the **Poisson distribution** is as an approximation to the binomial distribution when n is large and p is small.
- It is the case when n is large and p is small the mass function depends almost entirely on the mean np, and very little on the specific values of n and p.
- We can therefore approximate the binomial mass function with a quantity $\lambda = np$; this λ is the parameter in the Poisson distribution.

Poisson R.V.: pmf, mean, and variance

- If X ~ Poisson(λ), the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!}, & \text{for } x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- Mean: $\mu_X = \lambda$

- Variance:

$$\sigma_X^2 = \lambda$$

Note: X is a discrete random variable and λ must be a positive constant.

Example 8

Particles are suspended in a liquid medium at a concentration of 6 particles per mL. A large volume of the suspension is thoroughly agitated, and then 3 mL are withdrawn. What is the probability that exactly 15 particles are withdrawn?

Poisson Distribution to Estimate Rate

Let λ denote the mean number of events that occur in one unit of time or space. Let X denote the number of events that are observed to occur in t units of time or space.

If $X \sim$ Poisson(λt), we estimate λ with $\hat{\lambda} = \dfrac{X}{t}$.

Notes on Estimating a Rate

- $\hat{\lambda}$ is unbiased.
- The uncertainty in $\hat{\lambda}$ is $\sigma_{\hat{\lambda}} = \sqrt{\dfrac{\lambda}{t}}$.
- In practice, we substitute $\hat{\lambda}$ for λ, since λ is unknown.

Example 9

A 5 mL sample of a suspension is withdrawn, and 47 particles are counted. Estimate the mean number of particles per mL, and find the uncertainty in the estimate.

Section 4.4: Some Other Discrete Distributions

Hypergeometric Distribution:

- Consider a finite population containing two types of items, which may be called successes and failures.
- A simple random sample is drawn from the population.
- Each item sampled constitutes a Bernoulli trial.
- As each item is selected, the probability of successes in the remaining population decreases or increases, depending on whether the sampled item was a success or a failure.

Hypergeometric

- For this reason the trials are not independent, so the number of successes in the sample does not follow a binomial distribution.
- The distribution that properly describes the number of successes is the **hypergeometric distribution**.

Hypergeometric pmf

Assume a finite population contains N items, of which R are classified as successes and $N - R$ are classified as failures. Assume that n items are sampled from this population, and let X represent the number of successes in the sample. Then X has a hypergeometric distribution with parameters N, R, and n, which can be denoted $X \sim \mathrm{H}(N, R, n)$.

Hypergeometric pmf

The probability mass function of *X* is

$$p(x) = P(X = x) = \begin{cases} \dfrac{\dbinom{R}{x}\dbinom{N-R}{n-x}}{\dbinom{N}{n}}, & \max(0, R+n-N) \le x \le \min(n, R) \\ 0, & \text{otherwise} \end{cases}$$

Mean and Variance of the Hypergeometric Distribution

If $X \sim H(N, R, n)$, then

- Mean of X: $\mu_X = \dfrac{nR}{N}$

- Variance of X: $\sigma_X^2 = n\left(\dfrac{R}{N}\right)\left(1-\dfrac{R}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

Example 10

Of 50 buildings in an industrial park, 12 have electrical code violations. If 10 buildings are selected at random for inspection, what is the probability that exactly 3 of the 10 have code violations? What are the mean and variance of X?

Geometric Distribution

- Assume that a sequence of independent Bernoulli trials is conducted, each with the same probability of success, p.
- Let X represent the number of trials up to and including the first success.
- Then X is a discrete random variable, which is said to have the **geometric distribution** with parameter p.
- We write $X \sim \text{Geom}(p)$.

Geometric R.V.: pmf, mean, and variance

If $X \sim$ Geom(p), then

- The pmf of X is

$$p(x) = P(X = x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- The mean of X is $\mu_X = \dfrac{1}{p}$.

- The variance of X is $\sigma_X^2 = \dfrac{1-p}{p^2}$.

Example 11

A test of weld strength involves loading welded joints until a fracture occurs. For a certain type of weld, 80% of the fractures occur in the weld itself, while the other 20% occur in the beam. A number of welds are tested. Let X be the number of tests up to and including the first test that results in a beam fracture.

1. What is the distribution of X?
2. Find $P(X = 3)$.
3. What are the mean and variance of X?

Negative Binomial Distribution

The negative binomial distribution is an extension of the geometric distribution. Let r be a positive integer. Assume that independent Bernoulli trials, each with success probability p, are conducted, and let X denote the number of trials up to and including the r^{th} success. Then X has the **negative binomial distribution** with parameters r and p. We write $X \sim \text{NB}(r,p)$.

Note: If $X \sim \text{NB}(r,p)$, then $X = Y_1 + \ldots + Y_r$ where $Y_1,\ldots,Y_r$ are independent random variables, each with Geom(p) distribution.

Negative Binomial R.V.: pmf, mean, and variance

If $X \sim NB(r,p)$, then

- The pmf of X is

$$p(x) = P(X = x) = \begin{cases} \binom{x-1}{r-1} p^{r}(1-p)^{x-r}, & x = r, r+1, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- The mean of X is $\mu_X = \dfrac{r}{p}$.
- The variance of X is $\sigma_X^2 = \dfrac{r(1-p)}{p^2}$.

Example 11 cont.

Find the mean and variance of X, where X represents the number of tests up to and including the third beam fracture.

Multinomial Trials

A Bernoulli trial is a process that results in one of two possible outcomes. A generalization of the Bernoulli trial is the **multinomial trial**, which is a process that can result in any of k outcomes, where $k \geq 2$.

We denote the probabilities of the k outcomes by $p_1, \ldots, p_k$.

Multinomial Distribution

- Now assume that n independent multinomial trials are conducted each with k possible outcomes and with the same probabilities $p_1,\ldots,p_k$.
- Number the outcomes 1, 2, …, k. For each outcome i, let X_i denote the number of trials that result in that outcome.
- Then $X_1,\ldots, X_k$ are discrete random variables.
- The collection $X_1,\ldots ,X_k$ said to have the **multinomial distribution** with parameters n, $p_1,\ldots, p_k$. We write $X_1,\ldots, X_k \sim \text{MN}(n, p_1,\ldots, p_k)$.

Multinomial R.V.

If $X_1,\ldots, X_k \sim MN(n, p_1,\ldots, p_k)$, then the pmf of $X_1,\ldots, X_k$ is

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x_1!\,x_2!\cdots x_k!}\, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \quad x_i = 0,1,2,\ldots,k \\ \qquad\qquad\qquad \text{and } \sum x_i = n \\ 0, \qquad \text{otherwise} \end{cases}$$

Note that if $X_1,\ldots, X_k \sim MN(n, p_1,\ldots, p_k)$, then for each i,
$X_i \sim Bin(n, p_i)$.

Example 12

The items produced on an assembly line are inspected, and each is classified as either conforming (acceptable), downgraded, or rejected. Overall, 70% of the items are conforming, 20% are downgraded, and 10% are rejected. Assume that four items are chosen independently and at random. Let X_1, X_2, X_3 denote the numbers among the 4 that are conforming, downgraded, and rejected, respectively.

1. What is the distribution of X_1, X_2, X_3?
2. What is the probability that 3 are conforming and 1 is rejected in a given sample?

Section 4.5: The Normal Distribution

The **normal distribution** (also called the Gaussian distribution) is by far the most commonly used distribution in statistics. This distribution provides a good model for many, although not all, continuous populations.

The normal distribution is continuous rather than discrete. The mean of a normal population may have any value, and the variance may have any positive value.

Normal R.V.: pdf, mean, and variance

The probability density function of a normal population with mean μ and variance σ^2 is given by

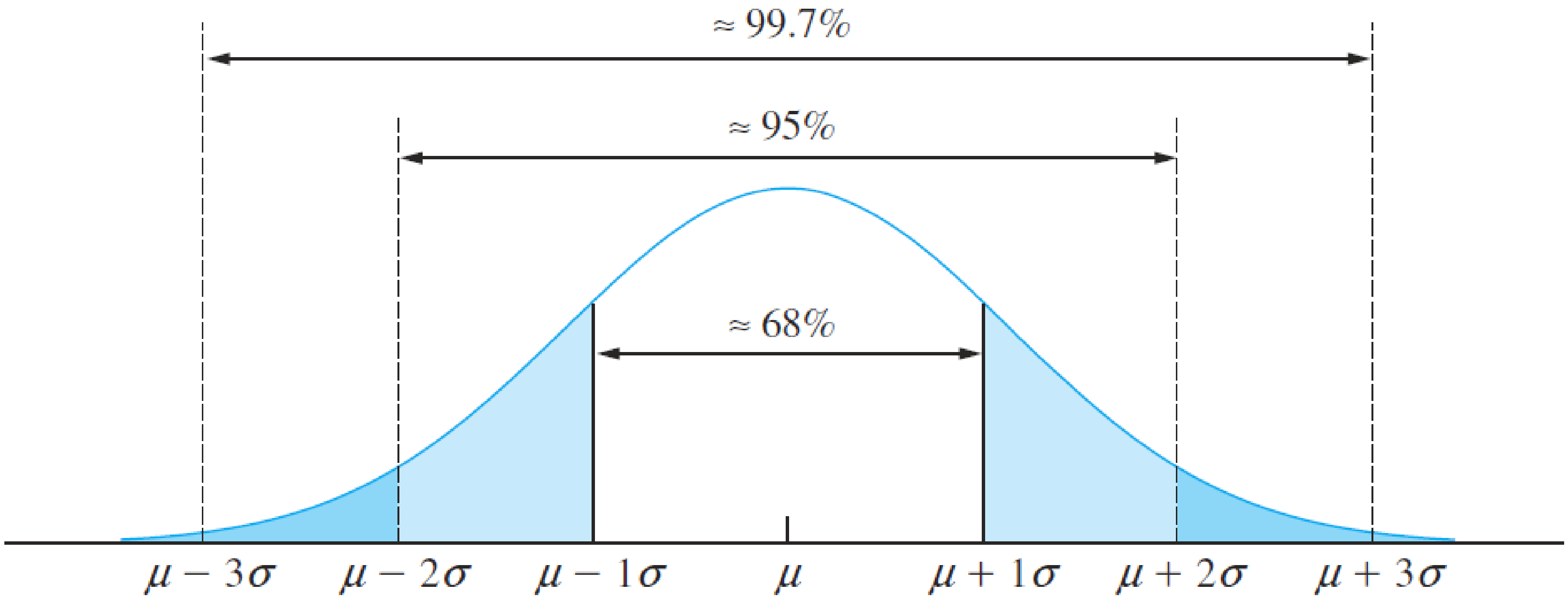
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

If $X \sim N(\mu, \sigma^2)$, then the mean and variance of X are given by

$$\mu_X = \mu$$

$$\sigma_X^2 = \sigma^2$$

68-95-99.7% Rule



This figure represents a plot of the normal probability density function with mean μ and standard deviation σ. Note that the curve is symmetric about μ, so that μ is the median as well as the mean. It is also the case for the normal population.

- About 68% of the population is in the interval $\mu \pm \sigma$.
- About 95% of the population is in the interval $\mu \pm 2\sigma$.
- About 99.7% of the population is in the interval $\mu \pm 3\sigma$.

Standard Units

- The proportion of a normal population that is within a given number of standard deviations of the mean is the same for any normal population.
- For this reason, when dealing with normal populations, we often convert from the units in which the population items were originally measured to **standard units**.
- Standard units tell how many standard deviations an observation is from the population mean.

Standard Normal Distribution

In general, we convert to standard units by subtracting the mean and dividing by the standard deviation. Thus, if x is an item sampled from a normal population with mean μ and variance σ^2, the standard unit equivalent of x is the number z, where

$$z = \frac{x - \mu}{\sigma}$$

The number z is sometimes called the "z-score" of x. The z-score is an item sampled from a normal population with mean 0 and standard deviation of 1. This normal population is called the **standard normal population**.

Example 13

Aluminum sheets used to make beverage cans have thicknesses (in thousandths of an inch) that are normally distributed with mean 10 and standard deviation 1.3. A particular sheet is 10.8 thousandths of an inch thick. Find the *z*-score.

Example 13 cont.

The thickness of a certain sheet has a z-score of -1.7. Find the thickness of the sheet in the original units of thousandths of inches.

Finding Areas Under the Normal Curve

The proportion of a normal population that lies within a given interval is equal to the area under the normal probability density above that interval. This would suggest integrating the normal pdf, but this integral does not have a closed form solution.

Finding Areas Under the Normal Curve

So, the areas under the standard normal curve (mean 0, variance 1) are approximated numerically and are available in a **standard normal table** or **z table**, given in Table A.2.

We can convert any normal into a standard normal so that we can compute areas under the curve.

The table gives the area in the left-hand tail of the curve. Other areas can be calculated by subtraction or by using the fact that the total area under the curve is 1.

Example 14

Find the area under normal curve to the left of $z = 0.47$.

Find the area under the curve to the right of $z = 1.38$.

Example 15

Find the area under the normal curve between $z = 0.71$ and $z = 1.28$.

What z-score corresponds to the 75^{th} percentile of a normal curve?

Estimating the Parameters

If $X_1, \ldots, X_n$ are a random sample from a $N(\mu, \sigma^2)$ distribution, μ is estimated with the sample mean and σ^2 is estimated with the sample standard deviation.

As with any sample mean, the uncertainty in $\bar{X}$ is $\sigma / \sqrt{n}$ which we replace with $s / \sqrt{n}$, if σ is unknown. The mean is an unbiased estimator of μ.

Linear Functions of Normal Random Variables

Let $X \sim N(\mu, \sigma^2)$ and let $a \neq 0$ and b be constants.

Then

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

Let $X_1, X_2, \ldots, X_n$ be independent and normally distributed with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$.

Let $c_1, c_2, \ldots, c_n$ be constants, and $c_1 X_1 + c_2 X_2 + \ldots + c_n X_n$ be a linear combination. Then

$$c_1 X_1 + c_2 X_2 + \cdots + c_n X_n \sim$$

$$N(c_1\mu_1 + c_2\mu_2 + \cdots + c_n\mu_n, c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_n^2\sigma_n^2)$$

Example 16

A chemist measures the temperature of a solution in °C. The measurement is denoted C, and is normally distributed with mean 40°C and standard deviation 1°C. The measurement is converted to °F by the equation $F = 1.8C + 32$. What is the distribution of F?

Distributions of Functions of Normals

Let $X_1, X_2, \ldots, X_n$ be independent and normally distributed with mean μ and variance σ^2. Then

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Let X and Y be independent, with $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Then

$$X + Y \sim N(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2)$$

$$X - Y \sim N(\mu_X - \mu_Y,\ \sigma_X^2 + \sigma_Y^2)$$

Section 4.6: The Lognormal Distribution

For data that contain outliers, the normal distribution is generally not appropriate. The **lognormal distribution**, which is related to the normal distribution, is often a good choice for these data sets.

If $X \sim N(\mu, \sigma^2)$, then the random variable $Y = e^X$ has the lognormal distribution with parameters μ and σ^2.

If Y has the lognormal distribution with parameters μ and σ^2, then the random variable $X = \ln Y$ has the $N(\mu, \sigma^2)$ distribution.

Lognormal pdf, mean, and variance

The pdf of a lognormal random variable with parameters μ and σ^2 is

$$f(x) = \begin{cases} \dfrac{1}{\sigma x\sqrt{2\pi}} \exp\left[-\dfrac{1}{2\sigma^2}(\ln x - \mu)^2\right], x > 0 \\ 0, \quad \text{otherwise} \end{cases}$$

Section 4.7:
The Exponential Distribution

The **exponential distribution** is a continuous distribution that is sometimes used to model the time that elapses before an event occurs. Such a time is often called a waiting time.

The probability density of the exponential distribution involves a parameter, which is a positive constant λ whose value determines the density function's location and shape.

We write $X \sim \text{Exp}(\lambda)$.

Exponential R.V.: pdf, cdf, mean and variance

The pdf of an exponential r.v. is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

The cdf of an exponential r.v. is

$$F(x) = \begin{cases} 0, & x \le 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}.$$

Exponential R.V.: pdf, cdf, mean and variance

The mean of an exponential r.v. is

$$\mu_X = 1/\lambda.$$

The variance of an exponential r.v. is

$$\sigma_X^2 = 1/\lambda^2.$$

The Exponential Distribution and the Poisson Process

If events follow a Poisson process with rate parameter λ, and if T represents the waiting time from any starting point until the next event, then $T \sim \text{Exp}(\lambda)$.

Example 17

A radioactive mass emits particles according to a Poisson process at a mean rate of 15 particles per minute. At some point, a clock is started.

1. What is the probability that more than 5 seconds will elapse before the next emission?
2. What is the mean waiting time until the next particle is emitted?

Lack of Memory Property

The exponential distribution has a property known as the lack of memory property: If $T \sim \text{Exp}(\lambda)$, and t and s are positive numbers, then $P(T > t + s \mid T > s) = P(T > t)$.

If $X_1, \ldots, X_n$ are a random sample from $\text{Exp}(\lambda)$, then the parameter λ is estimated with

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

This estimator is biased. This bias is approximately equal to λ / n. The uncertainty in $\hat{\lambda}$ is estimated with

$$\sigma_{\hat{\lambda}} \approx \frac{1}{\bar{X}\sqrt{n}}$$

This uncertainty estimate is reasonably good when the sample size is more than 20.

Example 18

The number of hits on a website follows a Poisson process with a rate of 3 per minute.

1. What is the probability that more than a minute goes by without a hit?

2. If 2 minutes have gone by without a hit, what is the probability that a hit will occur in the next minute?

Section 4.8: The Uniform, Gamma and Weibull Distributions

The **uniform distribution** has two parameters, *a* and *b*, with $a < b$. If X is a random variable with the continuous uniform distribution then it is uniformly distributed on the interval (a, b). We write $X \sim U(a, b)$.

The pdf is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

Mean and Variance

Let $X \sim U(a, b)$.

Then the mean is

$$\mu_X = \frac{a+b}{2}$$

and the variance is

$$\sigma_X^2 = \frac{(b-a)^2}{12}.$$

Example 19

When a motorist stops at a red light at a certain intersection, the waiting time for the light to turn green, in seconds, is uniformly distributed on the interval (0, 30). Find the probability that the waiting time is between 10 and 15 seconds.

The Gamma Distribution

First, let's consider the gamma function:

For $r > 0$, the **gamma function** is defined by

$$\Gamma(r) = \int_0^{\infty} t^{r-1} e^{-t} dt \quad .$$

The gamma function has the following properties:

1. If r is any integer, then $\Gamma(r) = (r-1)!$.
2. For any r, $\Gamma(r+1) = r\,\Gamma(r)$.
3. $\Gamma(1/2) = \sqrt{\pi}$.

Gamma R.V.

The pdf of the gamma distribution with parameters $r > 0$ and $\lambda > 0$ is

$$f(x) = \begin{cases} \dfrac{\lambda x^{r-1} e^{-\lambda x}}{\Gamma(r)}, x > 0 \\ 0, \quad x \le 0 \end{cases}.$$

The mean and variance are given by

$\mu_X = r/\lambda$ and $\sigma_X^2 = r/\lambda^2$, respectively.

Gamma R.V.

If $X_1, \ldots, X_r$ are independent random variables, each distributed as Exp(λ), then the sum $X_1 + \ldots + X_r$ is distributed as a gamma random variable with parameters r and λ, denoted as $\Gamma(r, \lambda)$.

Example 20

Assume that arrival times at a drive-through window follow a Poisson process with mean $\lambda = 0.2$ arrivals per minute. Let T be the waiting time until the third arrival.

Find the mean and variance of T.

Find $P(T \leq 20)$.

The Weibull Distribution

The Weibull distribution is a continuous random variable that is used in a variety of situations. A common application of the Weibull distribution is to model the lifetimes of components. The Weibull probability density function has two parameters, both positive constants, that determine the location and shape. We denote these parameters α and β.

If $\alpha = 1$, the Weibull distribution is the same as the exponential distribution with parameter $\lambda = \beta$.

Weibull R.V.

The pdf of the Weibull distribution is

$$f(x)=\begin{cases}\alpha\beta^{\alpha}x^{\alpha-1}e^{-(\beta x)^{\alpha}}, x>0\\ 0, \quad x\le 0\end{cases}.$$

The mean of the Weibull is

$$\mu_X=\frac{1}{\beta}\Gamma\left(1+\frac{1}{\alpha}\right).$$

The variance of the Weibull is

$$\sigma_X^2=\frac{1}{\beta^2}\left\{\Gamma\left(1+\frac{2}{\alpha}\right)-\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2\right\}.$$

Section 4.9: Some Principles of Point Estimation

- We collect data for the purpose of estimating some numerical characteristic of the population from which they come.
- A quantity calculated from the data is called a statistic, and a statistic that is used to estimate an unknown constant, or parameter, is called a point estimator. Once the data has been collected, we call it a point estimate.

Questions of Interest

- Given a point estimator, how do we determine how good it is?
- What methods can be used to construct good point estimators?

Notation: θ is used to denote an unknown parameter, and $\hat{\theta}$ to denote an estimator of θ.

Measuring the Goodness of an Estimator

- The accuracy of an estimator is measured by its bias, and the precision is measured by its standard deviation, or uncertainty.
- To measure the overall goodness of an estimator, we used the mean squared error (MSE) which combines both bias and uncertainty.

Mean Squared Error

Let θ be a parameter, and $\hat{\theta}$ an estimator of θ. The mean squared error (MSE) of $\hat{\theta}$ is

$$\text{MSE}_{\hat{\theta}} = (\mu_{\hat{\theta}} - \theta)^2 + \sigma_{\hat{\theta}}^2$$

An equivalent expression for the MSE is

$$\text{MSE}_{\hat{\theta}} = \mu_{(\hat{\theta}-\theta)^2}$$

Example 21

Let $X \sim \text{Bin}(n, p)$ where p is unknown. Find the MSE of the estimator $\hat{p} = X / n$.

Method of Maximum Likelihood

- The idea is to estimate a parameter with the value that makes the observed data most likely.
- When a probability mass function or probability density function is considered to be a function of the parameters, it is called a **likelihood function**.
- The **maximum likelihood estimate** is the value of the estimators that when substituted in for the parameters maximizes the likelihood function.

Desirable Properties

Maximum likelihood is the most commonly used method of estimation. The main reason for this is that in most cases that arise in practice, MLEs have two very desirable properties,

1. In most cases, as the sample size n increases, the bias of the MLE converges to 0.
2. In most cases, as the sample size n increases, the variance of the MLE converges to a theoretical minimum.

Section 4.10: Probability Plots

Scientists and engineers often work with data that can be thought of as a random sample from some population. In many cases, it is important to determine the probability distribution that approximately describes the population.

More often than not, the only way to determine an appropriate distribution is to examine the sample to find a probability distribution that fits.

Finding a Distribution

- Probability plots are a good way to determine an appropriate distribution.
- Here is the idea: Suppose we have a random sample $X_1,\ldots, X_n$. We first arrange the data in ascending order. Then assign evenly spaced values between 0 and 1 to each X_i. There are several acceptable ways to this; the simplest is to assign the value $(i - 0.5)/n$ to X_i .

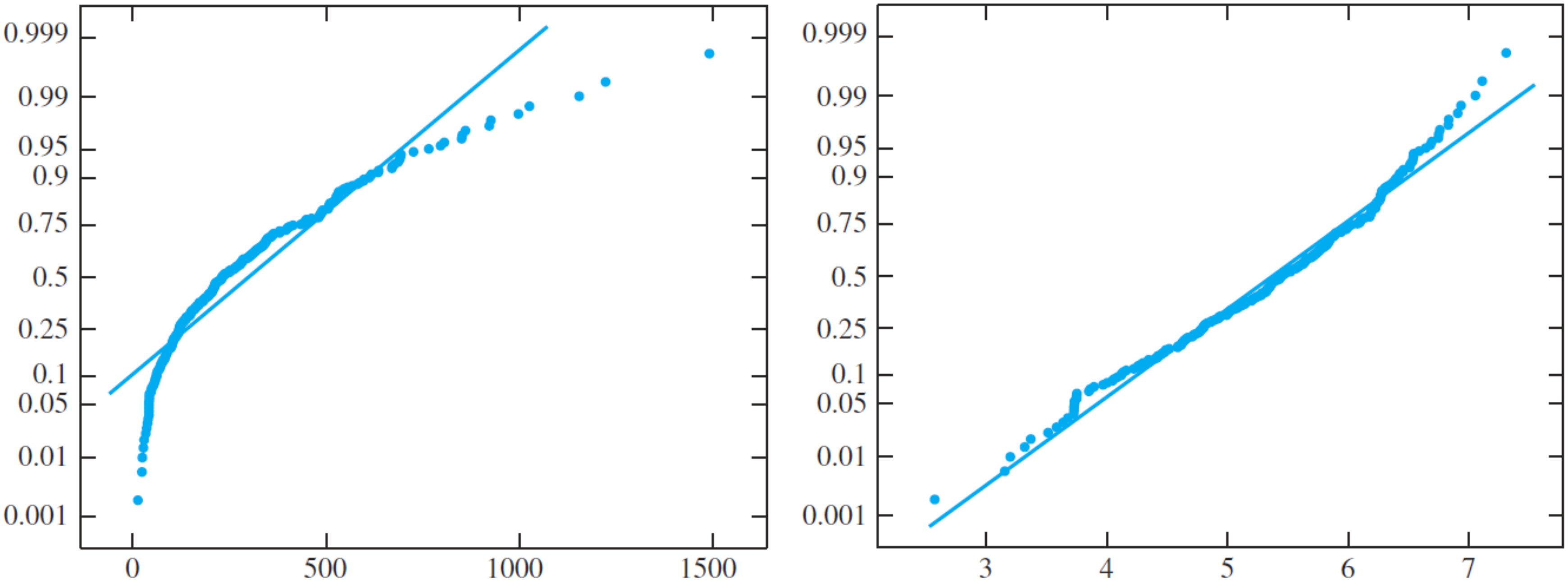
The distribution that we are comparing the X's to should have a mean and variance that match the sample mean and variance. We want to plot

$(X_i, F(X_i))$, if this plot resembles the cdf of the distribution that we are interested in, then we conclude that that is the distribution the data came from.

Software

Many software packages take the $(i - 0.5)/n$ assigned to each X_i, and calculate the quantile (Q_i) corresponding to that number from the distribution of interest. Then it plots each (X_i, Q_i). If this plot is a reasonably straight line then you may conclude that the sample came from the distribution that we used to find quantiles.

Normal Probability Plots



The sample plotted on the left comes from a population that is not close to normal.

The sample plotted on the right comes from a population that is close to normal.

Section 4.11: The Central Limit Thereom

The Central Limit Theorem

Let $X_1,\ldots,X_n$ be a random sample from a population with mean μ and variance σ^2.

Let $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n}$ be the sample mean.

Let $S_n = X_1+\ldots+X_n$ be the sum of the sample observations. Then if n is sufficiently large,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{and} \quad S_n \sim N(n\mu, n\sigma^2)$$

approximately.

Rule of Thumb for the CLT

For most populations, if the sample size is greater than 30, the Central Limit Theorem approximation is good.

Two Examples of the CLT

Normal approximation to the Binomial:

If $X \sim \text{Bin}(n,p)$ and if $np > 10$, and $n(1-p) > 10$, then

$X \sim N(np, np(1-p))$ approximately and

$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ approximately.

Normal Approximation to the Poisson:

If $X \sim \text{Poisson}(\lambda)$, where $\lambda > 10$, then $X \sim N(\lambda, \lambda^2)$.

Continuity Correction

- The binomial distribution is discrete, while the normal distribution is continuous.
- The continuity correction is an adjustment, made when approximating a discrete distribution with a continuous one, that can improve the accuracy of the approximation.
- If you want to include the endpoints in your probability calculation, then extend each endpoint by 0.5. Then proceed with the calculation.
- If you want exclude the endpoints in your probability calculation, then include 0.5 less from each endpoint in the calculation.

Example 22

The manufacturer of a certain part requires two different machine operations. The time on machine 1 has mean 0.4 hours and standard deviation 0.1 hours. The time on machine 2 has mean 0.45 hours and standard deviation 0.15 hours. The times needed on the machines are independent. Suppose that 65 parts are manufactured. What is the distribution of the total time on machine 1? On machine 2? What is the probability that the total time used by both machines together is between 50 and 55 hours?

Example 23

If a fair coin is tossed 100 times, use the normal curve to approximate the probability that the number of heads is between 45 and 55 *inclusive*.

Section 4.12: Simulation

Simulation refers to the process of generating random numbers and treating them as if they were data generated by an actual scientific distribution. The data so generated are called **simulated** or **synthetic** data.

Example 24

An engineer has to choose between two types of cooling fans to install in a computer. The lifetimes, in months, of fans of type A are exponentially distributed with mean 50 months, and the lifetime of fans of type B are exponentially distributed with mean 30 months. Since type A fans are more expensive, the engineer decides that she will choose type A fans if the probability that a type A fan will last more than twice as long as a type B fan is greater than 0.5. Estimate this probability.

Simulation

- We perform a simulation experiment, using samples of size 1000.
- Generate a random sample $A_1^*, A_2^*, \ldots, A_{1000}^*$ from an exponential distribution with mean 50 (λ = 0.02).
- Generate a random sample $B_1^*, B_2^*, \ldots, B_{1000}^*$ from an exponential distribution with mean 30 (λ = 0.033).
- Count the number of times that $A_i^* > 2B_i^*$.
- Divide the number of times that $A_i^* > 2B_i^*$ occurred by the total number of trials. This is the estimate of the probability that type A fans last twice as long as type B fans.

Summary

- We considered various discrete distributions: Bernoulli, Binomial, Poisson, Hypergeometric, Geometric, Negative Binomial, and Multinomial.
- Then we looked at some continuous distributions: Normal, Exponential, Uniform, Gamma, and Weibull.
- We learned about the Central Limit Theorem.
- We discussed Normal approximations to the Binomial and Poisson distributions.
- Finally, we discussed simulation studies.