CS 660: Mathematical Foundations of Analytics

Dr. Francis Parisi

Pace University

Spring 2018

IMS Chapter 3 – Special Distributions
FPA Sections 2.1

Special Distributions & Their Properties

In this section we will discuss some common distributions that we will use in our study of statistical inference and models

Special Distributions & Their Properties

Uniform Distribution

The Uniform distribution is the simplest distribution over the interval zero to one, denoted $U(0,1)$. It's pdf is

$$u(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

and we note that this is a density function.

In general if $X \sim U(\alpha, \beta)$ its density is given by

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

Uniform random variables are useful for generating other random variables on a computer

Fundamental Distributions & Their Properties

Example

If X is uniformly distributed over the interval $(0, 10)$, what is the probability that

1. $X < 3$
2. $X > 5$
3. $3 < X < 9$?

Special Distributions & Their Properties

Exercise

Buses arrive at a specified stop at 15-minute intervals starting at 7 A.M. That is, they arrive at 7:00, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7:00 and 7:30, find the probability that

(a) she waits less than 5 minutes for a bus

(b) more than 10 minutes for a bus

- A **Bernoulli experiment** is a random experiment with a pair of mutually exclusive and exhaustive outcomes
- female or male, yes or no, defective or non-defective
- When a Bernoulli experiment is repeated many independent times we have a sequence of **Bernoulli trials**
- Let X be a random variable associated with a Bernoulli trial defined as

$$X(\text{success}) = 1 \text{ and } X(\text{failure}) = 0$$

- The pmf of X is

$$p(x) = p^x(1-p)^{1-x}$$

  and X has a **Bernoulli distribution**

Special Distributions & Their Properties

For a Bernoulli random variable X we have the expected value of X

$$\mu = \mathbb{E}[X] = (0)(1-p) + (1)(p) = p$$

and the variance

$$\sigma^2 = \mathrm{Var}\left(X\right) = p^2(1-p) + (1-p)^2 p = p(1-p)$$

and the standard deviation

$$\sigma = \sqrt{p(1-p)}$$

Special Distributions & Their Properties

- A sequence of Bernoulli trials is an n-tuple of zeros and ones
- We're often interested in the total number of 'successes' without regard to order
- If X is the possible number of success in n trials
- If x successes occur then we have $n - x$ 'failures'
- The number of ways the x successes can occur in n trials is

$$\binom{n}{x} = \frac{n!}{x!(n - x!)}$$

- The probability of success is p and the probability of failure is $1 - p$

Special Distributions & Their Properties

This leads to the pmf

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, \ldots, n$$

This is the pmf of the **binomial distribution**

Special Distributions & Their Properties

Example

Roll a fair six-sided die three times What is the probability of getting exactly two sixes?

Solution

X has a binomial distribution with $n = 3$, and $p = 1/6$ so we have

$$P\left[X = 2\right] = p(2) = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = 0.06944$$

While hand calculation is easy when n and x are small, they become cumbersome as these values get large

Special Distributions & Their Properties

We can easily compute this as follows

```
> dbinom(2, 3, 1/6)
[1] 0.0694444

> dbinom(12, 30, 1/6)
[1] 0.00149247

In [1]: from scipy.stats import binom
In [2]: binom.pmf(2, 3, 1/6)
Out[2]: 0.06944444

and

In [3]: binom.pmf(12, 30, 1/6)
Out[3]: 0.00149247
```

Special Distributions & Their Properties

Negative Binomial and Geometric Distributions

Let Y denote the total number of failures in a sequence of independent Bernoulli trials before the r^{th} success so $Y + r$ is the number of trials necessary to get r successes
Then the pmf og Y is

$$p_Y(y) = \binom{y+r-1}{r-1} p^r (1-p)^y$$

and Y has a **negative binomial distribution**

If $r = 1$ then the pmf of Y is

$$p_Y(y) = p(1-p)^y$$

and Y has a **geometric distribution** moreover Y is the number of failures until the first success

The binomial distribution generalizes to the **multinomial distribution**

Rolling a fair die is one of six categories $C_1, C_2, \ldots, C_6$ with probabilities $p(1) = p(2) = \cdots = p(6) = 1/6$

If we roll the die n times and X_i is the number of times C_i results we have the joint pmf

$$P\left[X_1 = x_1, X_2 = x_2, \ldots, X_5 = x_6\right] = \frac{n!}{x_1! \cdots x_5! x_6!} = p^{x_1} \cdots p^{x_5} p^{x_6}$$

In general we have

$$P\left[X_1 = x_1, X_2 = x_2, \ldots, X_{k-1} = x_{k-1}\right] = \frac{n!}{x_1! \cdots x_{k-1}! x_k!} = p^{x_1} \cdots p^{x_{k-1}} p^k$$

Special Distributions & Their Properties

Hypergeometric Distribution

- Suppose we have a lot of widgets of size N
- Of these, the number of defective items is D
- We draw a sample from this lot *without replacement* of n items
- The number of defects in this sample is X
- The probability that $X = x$ items are defective in our sample is

$$p(x) = \frac{\binom{N-D}{n-x}\binom{D}{x}}{\binom{N}{n}}, \quad \text{for } x = 0, 1, \ldots, n$$

- X has a **hypergeometric distribution**

Special Distributions & Their Properties

Poisson Distribution

The Poisson distribution comes into play when we are dealing with counting the number of events over a given time interval such as

- The number of tornadoes touching down in Kansas per year
- The number of customers entering a bank between 8:00 am and 10:00
- The number of car accidents in a busy intersection
- The number of hurricanes making landfall along the U.S. eastern coastline

Random variable X has a **Poisson Distribution** with pmf

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, 2, \ldots$$

Special Distributions & Their Properties

Example

Let X be the number of auto accidents at a busy intersection, and $X \sim \mathrm{Pois}(\lambda = 2)$
The probability of *at least* one accident in a week is

$$P\left[X \geq 1\right] = 1 - P\left[X = 0\right] = 1 - e^{-2}$$

```
> 1-dpois(0,2)
[1] 0.864665
```

```
In [3]: from scipy.stats import poisson
In [4]: 1-poisson.pmf(0,2)
Out[4]: 0.8646647
```

- The Γ **Distribution** with parameters (α, β) often arises, in practice as the distribution of the amount of time one has to wait until a total of n events has occurred
- The probability density function of the $\Gamma(\alpha, \beta)$ with $\alpha > 0, \beta > 0$ distribution is given by

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \qquad x > 0$$

- $\Gamma(\alpha)$ is the gamma function and is defined as

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

Special Distributions & Their Properties

- We evaluate $\Gamma(\alpha)$ using integration by parts to get

$$(\alpha - 1)\Gamma(\alpha - 1) \tag{3.1}$$

- For integral values of α, say $\alpha = n$ repeated application of (3.1) gives the well known

$$\Gamma(n) = (n - 1)!$$

- Finally, the mean and variance are

$$\mu = \mathbb{E}[X] = \alpha\beta \quad \text{and} \quad \sigma^2 = \mathrm{Var}\,(X) = \alpha\beta^2$$

Special Distributions & Their Properties

Example

Let X be the lifetime in hours of a certain battery, and X has a $\Gamma(5, 4)$ distribution. The mean lifetime is 20 hours and the standard deviation is $\sqrt{5 \times 16} = 8.94$ hours. the probability the battery lasts at least 50 hours is

```
> 1-pgamma(50, shape = 5, scale = 4)
[1] 0.00534551
```

and

```
In [15]: from scipy.stats import gamma
In [16]: 1-gamma.cdf(50, a=5, scale= 4)
Out[16]: 0.005345505
```

Example (continued)

The probability that the battery's lifetime is within one standard deviation is

```
> pgamma(20+8.94,shape=5,scale=4)-pgamma(20-8.94, shape=5,scale=4)
[1] 0.700473
```

and

```
In [23]:  gamma.cdf(20+8.94,a=5,scale=4) - gamma.cdf(20-8.94,a=5,scale=4)
Out[23]:  0.7004733126222387
```

The χ^2-**Distribution** is a special case of the gamma distribution when $\alpha = r/2$ and $\beta = 2$ for some positive integer r
If X has the pdf

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, \qquad x > 0$$

then X has a χ^2-distribution with r degrees of freedom

The gamma distribution has unbounded support, that is,
$0 < x < \infty$
If we consider the case where the support of X is bounded say on (a, b) then we have the **β-distribution** with pdf

$$f(x) = \frac{1}{\Gamma(\alpha+\beta)} x^{\alpha+\beta-1} e^{-x}$$

for $0 < x < \infty$

Special Distributions & Their Properties
Normal Distribution

- **Normal Distribution** is likely the most common distribution, also known as the Gaussian distribution and is the familiar bell-shaped curve
- Because of its central importance we discuss some fundamental properties and characteristics

Definition

We say a random variable X has a normal distribution if its pdf is

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

The parameters μ and σ^2 are the mean and variance of X respectively. We write $X \sim N(\mu, \sigma^2)$.

Special Distributions & Their Properties

When $\mu = 0$ and $\sigma^2 = 1$ we have the **standard normal distribution** with density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = P(X < x), \qquad -\infty < x < \infty$$

The cumulative standard normal distribution is given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

This will be used often in statistical inference

Special Distributions & Their Properties

- If $X \sim N(\mu, \sigma^2)$ then
$$Z = \frac{X - \mu}{\sigma}$$
is distributed $N(0, 1)$, and is a standard normal r.v.
- The values of the standard normal distribution are often found in tables at the back of texts labeled – "Area under a Normal Curve" – or with a computer

```
pnorm(1.96)
[1] 0.975002
```

```
In [1]: from scipy.stats import norm
In [2]: norm.cdf(1.96)
Out[5]: 0.975002
```

This is the cumulative probability (area) to the left of $Z = 1.96$

Example

An expert witness in a paternity suit testifies that the length (in days) of human gestation is approximately normally distributed with parameters $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had the very long or very short gestation indicated by the testimony?

Special Distributions & Their Properties

Solution

We need to find the probability that the length of the pregnancy X is either greater than 290 days or less than 240 days

$$
\begin{aligned}
P(X > 290 \textit{ or } X < 240) &= P(X > 290) + P(X < 240) \\
&= P\left(Z > \frac{290 - 270}{10}\right) + P\left(Z < \frac{240 - 270}{10}\right) \\
&= P\left(Z > 2\right) + P\left(Z < -3\right) \\
&= 1 - \Phi(2) + 1 - \Phi(3) \\
&\approx 0.0241
\end{aligned}
$$

```
> 1-pnorm(2) + 1-pnorm(3)
[1] 0.0241
```

```
In [1]: 1-norm.cdf(2) + 1-norm.cdf(3)
Out[1]: 0.024100
```

Special Distributions & Their Properties

In the prior example we made use of the following relationships between probabilities and the area under the normal curve $\Phi(x)$

$$
\begin{aligned}
P(Z < x) &= \Phi(x) \\
P(Z > x) &= 1 - \Phi(x) \\
P(Z < -x) &= 1 - \Phi(x) \\
P(Z > -x) &= 1 - (1 - \Phi(x)) = \Phi(x)
\end{aligned}
$$

The **empirical rule** is of practical importance as one often considers the probability that a random variable is within one, two or three standard deviations of the mean

- $P\left[\mu - \sigma < X < \mu + \sigma\right] \approx 0.6827$
- $P\left[\mu - 2\sigma < X < \mu + 2\sigma\right] \approx 0.9545$
- $P\left[\mu - 3\sigma < X < \mu + 3\sigma\right] \approx 0.9973$

The empirical rule states that about 68% of the area is within one standard deviation, about 95% of the area is within two standard deviations, and about 99% is within three standard deviations

The normal distribution can be extended to cover jointly distributed normal random variables

We'll look at the **bivariate normal distribution**

Definition

A random vector X, Y has a bivariate normal distribution if its pdf is

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-q/2}, -\infty < x < \infty, -\infty < y < \infty$$

where

$$q = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Special Distributions & Their Properties

We can extend this to n-dimensions. Let $\mathbf{X} = (X_1, \ldots, X_n)'$ and the $X_i \sim N(\mu, \sigma^2)$. Then the pdf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\Big\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big\}, \text{ for } \mathbf{x} \in \mathbb{R}$$

Definition

Suppose $\varphi(y)$ is the normal distribution $N(\mu, \sigma^2)$, and Y has a normal distribution. If $Y = \log X$ then x has a **lognormal distribution.**

To derive the probability density function of the lognormal distribution we proceed as follows:

$$\varphi(y)dy = \varphi(\log x)d(\log x) = \frac{\varphi(\log x)}{x}dx = f(x)dx$$

therefore the pdf of the lognormal distribution, $LN(\mu, \sigma^2)$, is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}x}e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$$

Special Distributions & Their Properties

Exponential Distribution

The probability density function of the exponential distribution with a parameter $\lambda > 0$ is

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}$$

for $x \geq 0$.
Its mean and variance are

$$\mu = \mathbb{E}[x] = \lambda \text{ and } \sigma^2 = \operatorname{Var}(x) = \lambda^2$$

Note sometimes the exponential is given with parameter λ, in which case the mean and variance are $1/\lambda$ and $1/\lambda^2$ respectively

Special Distributions & Their Properties

Example

The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \frac{1}{100}e^{-x/100}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

What is the probability that

1. a computer will function between 50 and 150 hours before breaking down?
2. it will function for fewer than 100 hours?

Suppose we have two independent random variables, $W \sim N(01, 1)$ and $V \sim \chi^2(r)$. If we define a new random variable as

$$T = \frac{W}{\sqrt{V/r}}$$

then we say T has a ***t*-distribution**

The derivation is in the text: we first find the joint pdf which, because of independence, is simply the product of the two pdfs, define a transformation and from the joint pdf of the transformation we find the marginal pdf of t

Special Distributions & Their Properties

The t-Distribution plays a central role in hypothesis testing, as we'll see later in the course

If we wish to find $P[T < 2]$ where $T \sim t(15)$, T is a t-distributed random variable with 15 degrees of freedom we have

```
> pt(2,15)
[1] 0.968027
```

```
In [15]: from scipy.stats import t
In [16]: t.cdf(2,15)
Out[16]: 0.96802749635764
```

Consider two χ^2 random variables U and V with r_1 and r_2 degrees of freedom, respectively

Then the random variable

$$W = \frac{U/r_1}{V/r_2}$$

has an F-distribution with with r_1 and r_2 degrees of freedom

References

Hogg, R. V., McKean, J. W. and Craig, A. T. (2018).
Introduction to Mathematical Statistics.
Pearson, Boston, eighth edition.

Ross, S. (2010).
A First Course in Probability.
Prentice Hall, Upper Saddle River, eigth edition.

Wu, J. and Coggeshall, S. (2012).
Foundations of Predictive Analytics.
CRC Press, Boca Raton, FL.