Lecture 5: Continuous Random Variables

MATH 697

Omidali Jazi

September 18, 2018

McGill University

Goals for this Chapter

- Learn about continuous random variables:
  - Probability distributions
  - Expected value, variance and standard deviation
- Define specific distributions
  - Uniform
  - Normal
  - Gamma Family (Exponential, Chi-square)
  - Beta
- Define and apply Tchebycheff's Theorem
- Moments and moment-generating Functions

Continuous Random Variables

- Continuous random variables can take on any value within a range.
  Continuous random variables have an uncountably infinite number of values they can take on.
- Example: with a perfectly precise measuring instrument, blood pressure is continuous.
  - Compared to discrete data that can take on either a finite or countably infinite number of values

Continuous Random Variables

An important mathematical distinction with continuous random variables is that.

- For a continuous r.v., the probability that any particular value y occurs is always zero. i.e, for every y $P(Y = y) = 0$.
- This requires both a change in how we think about continuous r.v.s and a change in notation.
  - But, while the differences are important, the intuition that we've developed about random variables will carry over.

Some Definitions

Definition 1

Let Y denote any r. v. The Cumulative Distribution Function (CDF) of Y, denoted by $F(y)$, is defined by

$$F(y) = P(Y \leq y)$$

for $-\infty < y < \infty$.

CDFs for Discrete Random Variables

- Note that this is the same definition we used for discrete random variables.
- But some of the CDF's characteristics differ between discrete and continuous r.v.s.
- CDFs for discrete random variables are step functions They jump at each y value at which there is positive probability. Otherwise, in between, the function is flat.
- This occurs because the CDF only increases at the finite or countable number of points with positive probability.
- Also, the function is always a monotonic, nondecreasing function.

Properties of a Distribution Function

Theorem 2

If $F(y)$ is a CDF, then

1. $F(-\infty) = \lim_{y \to -\infty} P(Y \leq y) = \lim_{y \to -\infty} F(y) = 0$
2. $F(\infty) = \lim_{y \to \infty} P(Y \leq y) = \lim_{y \to \infty} F(y) = 1$
3. $F(y)$ *is a right-continuous non-decreasing function of* y.

Non-decreasing means that if y_1 and y_2 are any values such that $y_1 < y_2$ then $F(y_1) \leq F(y_2)$.

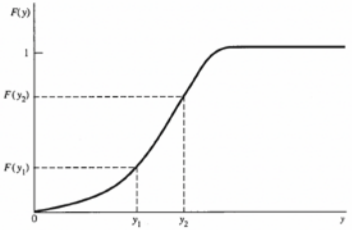
Distribution Functions for Continuous R.V.s

Definition 3

A random variable Y with the CDF $F(y)$ is said to be continuous if $F(y)$ is continuous for $-\infty < y < \infty$.

What this means is that the distribution function for continuous random variables is a smooth function.

FIGURE 4.2 Distribution function for a continuous random variable



Probability Density Functions

Definition 4

Let $F(y)$ be the distribution function for a continuous random variable Y. Then, $f(y)$ is given by

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

- Wherever the derivative exists, is called the probability density function (pdf) for the random variable Y.
- It is the analog of the probability mass function for discrete random variables.

A Note on the Notation

- The notation $f(y)$ is shorthand notation for the pdf of random variable Y evaluated at y.
  - If we want to be explicit, we can write $f_Y(y)$.
- With one random variable, it is usually clear without the subscript, but with two or more it can be confusing, unclear.
  - I.e., Let random variable Y have pdf $f_Y(.)$ and random variable X have pdf $f_X(.)$.
  - Also, sometimes different letters used to distinguish. Let random variable Y have pdf $f(.)$ and random variable X have pdf $g(.)$.

Properties of Density Functions

Theorem 5

If $f(y)$ is a pdf for a continuous random variable, then

1. $f(y) \geq 0$ *for all* y, $-\infty < y < \infty$.
2. $\int_{-\infty}^{\infty} f(t)dt = 1$

That is, pdfs are always non-negative and they must integrate to 1. Functions that do not have these properties cannot be density functions.

The Connection Between PDFs and CDFs

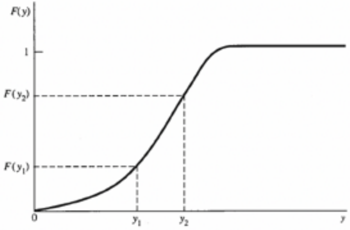
- It follows from the Definitions that y

$$F(y) = \int_{-\infty}^{y} f(t)dt$$

where $f(\cdot)$ is the pdf and t is the variable of integration.

- In a picture

FIGURE 4.2
Distribution function for a continuous random variable



Quantiles and Percentiles

Definition 6

For $0 < p < 1$

- If Y is continuous, the pth quantile of Y, denoted ϕ_p is the value such that $P(Y \leq \phi_p) = F(\phi_p) = p$.

- If Y is discrete, the pth colorblue quantile of Y, ϕ_p is the smallest value such that $P(Y \leq p) = F(\phi_p) \geq p$.

- $100 \times \phi_p$ is the 100pth percentile of Y.

- Note that $\phi_{0.5}$ is the median. $P(Y \leq \phi_{0.5}) = 0.5$.
- Other special percentiles:
  - Minimum: 0th percentile.
  - Maximum: 100th percentile.

Calculating Probabilities for Continuous R.V.s

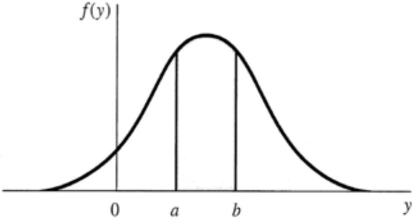
Theorem 7

If a random variable Y has density function $f(y)$ and $a < b$, then the probability that Y falls in the interval $[a, b]$ is

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

- Graphically: The area under the curve between a and b.

FIGURE **4.8**
$P(a \leq Y \leq b)$



A Couple of Notes

- Remember that if Y is continuous, then $P(Y = a) = P(Y = b) = 0$
- Thus, it follows: $P(a \leq Y \leq b) = P(a < Y < b)$.
  - That is, whether the endpoints are included in the integration or not does not matter for continuous random variables.
  - This is not true for discrete random variables.

Example

Example 8

Given $f(y) = cy^2$, $0 \leq y \leq 2$.

Find c so that $f(y)$ is a valid pdf.

Find $P(1 < Y < 2)$.

Solution: From the propert of the density function

$$\int_0^2 cy^2 dy = 1$$

Hence, $c\left[\frac{y^3}{3}|_0^2\right] = 1$ which leads to $c = 3/8$.

$$P(1 < Y < 2) = \int_1^2 \frac{3}{8} y^2 dy = \frac{7}{8}$$

Expected Value, Variance, & Std. Deviation

- The expected value and variance of a continuous random variable. are intuitively the same as discrete random variables.
  - Expected value is the value you would get if you drew an infinite number of observations from the distribution and averaged them.
  - Variance and standard deviation measure how spread out the distribution is and thus how variable the data will be that come from the distribution.
- Where the definitions differ is in the mathematical details.

Defining Expected Value

Definition 9

Let Y be a continuous random variable; then the expected value of Y is

$$E\left[Y\right] = \int_{-\infty}^{\infty} y f(y) dy$$

Definition 10

Let $g(Y)$ be a function of Y; then the expected value of $g(Y)$ is

$$E\left[g(Y)\right] = \int_{-\infty}^{\infty} g(y) f(y) dy$$

provided that the integral exists

Expected Value Theorems

The expected value theorems we proved for discrete r.v.s carry over to continuous r.v.s.

Theorem 11

Let c be a constant and $g(Y)$ and $g_1(Y), ..., g_k(Y)$ be functions of a random variable Y. Then the following results hold:

1. $E(c) = c$
2. $E[cg(Y)] = cE[g(Y)]$
3. $E[g_1(Y) + ... + g_k(Y)] = E[g_1(Y)] + ... + E[g_k(Y)]$.

Defining Variance and Standard Deviation

The definitions are precisely the same as with discrete random variables, the only difference is in how they're calculated.

Definition 12

For a r.v. Y with $E(Y) = \mu$. The variance of Y is defined as

$$V(Y) = E[(Y - \mu)^2]$$

But since Y is continuous, the calculation is

$$\sigma^2 = \int_{-\infty}^{\infty} (y - \mu)^2 f(y) dy$$

Uniform Distribution

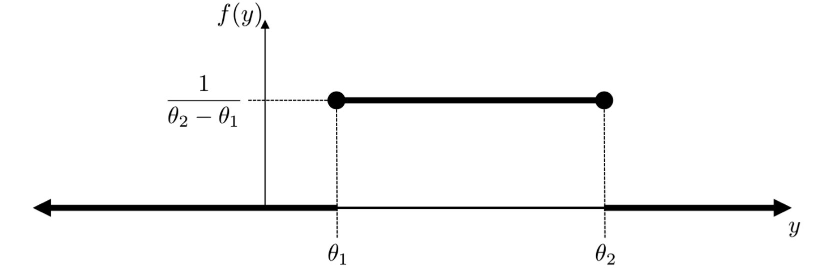
Definition 13

If $\theta_1 < \theta_2$, a random variable Y has a Uniform distribution on the interval (θ_1, θ_2), often denoted $U(\theta_1, \theta_2)$ if and only if the density of Y is

$$f(y) = \frac{1}{\theta_2 - \theta_1}, \qquad \theta_1 \leq y \leq \theta_2$$

Uniform Distribution

In a picture:



Uniform Distribution

- A Uniform distribution arises when the probability of an event occurring in some fixed interval is the same over all possible intervals of that size.
- Example: Imagine the example of bus arrival, where the bus is just as likely to arrive in any two-minute interval over a 10-minute window.

Why the Uniform Distribution?

- In experiments and surveys where you might need to select a random sample, the easiest way is to assign each entity a computer-generated uniformly distributed random number.
- Some continuous random variables in physical, biological, and other sciences have a uniform distribution.
- Example: If the number of arrivals into some system has a Poisson distribution and we are told that exactly one event happened in the interval $(0, t)$ then the time of occurrence of the event is uniformly distributed on $(0, t)$.

Expected Value and Variance

Theorem 14

If $Y \sim U(\theta_1, \theta_2)$, *then*

$$E(Y) = \frac{\theta_1 + \theta_2}{2}$$

and

$$V(Y) = \frac{(\theta_2 - \theta_1)^2}{12}$$

Defining Normal Distribution

Definition 15

A random variable Y is said to have a Normal distribution if and only if for $\sigma^2 > 0$ and $-\infty < \mu < \infty$, the density of Y is

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

The parameters for the normal distribution are μ and σ and it is denoted as $N(\mu, \sigma^2)$.

Theorem 16

If Y is normally distributed, then $E(Y) = \mu$ and $V(Y) = \sigma^2$

Normal Distribution

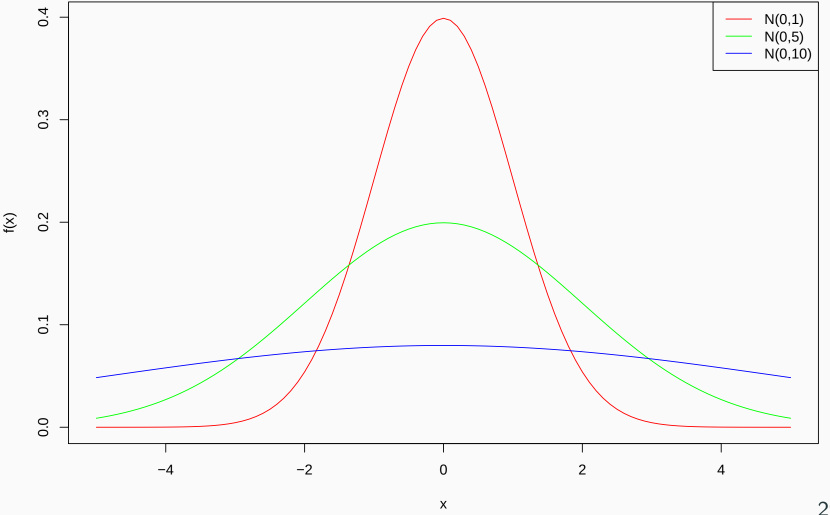
- The Normal distribution is widely used for a variety of reasons.
- Many natural phenomena are normally distributed.
- The asymptotic distibution of the sample mean is normal distibution (Central Limit Theorem). That is, it is useful for approximation.
- It is the famous “bell-shaped” distribution.
- The empirical rule we’ve been using is derived from it.

Rcode

```r
curve(dnorm(x, mean = 0, sd = 1), from = -5, to = 5, col = "red", ylab = "f(x)")
curve(dnorm(x, mean = 0, sd = 2), add = TRUE, col = "green")
curve(dnorm(x, mean = 0, sd = 5), add = TRUE, col = "blue")
legend("topright", legend = c("N(0,1)", "N(0,5)", "N(0,10)"),
       col = c("red","green","blue"), lty = 1)
```

Normal Distribution Curves

```
legend("topright", legend = c("N(0,1)", "N(0,5)", "N(0,10)"),
       col = c("red","green","blue"), lty = 1)
```



Standard Normal Distribution

- We use Z to represent a random variable from a standard normal distribution: $N(0,1)$
- If Z has a standard normal distribution, the cdf, $F(z) = P(Z \leq z)$, is often denoted by $\Phi(z)$
- Because the normal is symmetric for Z, for $\Phi(0) = 0.5$ and $\Phi(z) = 1 - \Phi(-z)$
- Also note that

$$P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z)$$

$$P(a < Z < b) = \Phi(b) - \Phi(a)$$

Standardizing Normal Distribution

- Standardizing means transforming an observation from a $N(\mu, \sigma^2)$ into a $N(0, 1)$ observation.
  If Y comes from a $N(\mu, \sigma^2)$ then $Z = (Y - \mu)/\sigma$ is $N(0, 1)$.

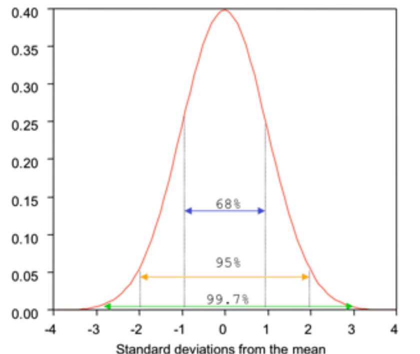
- Using the new notation:

$$\begin{aligned} P(a \leq Y \leq b) &= P(\frac{a-\mu}{\sigma} \leq Y \leq \frac{b-\mu}{\sigma}) \\ &= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}) \end{aligned}$$

The Empirical Rule

- For the normal distribution.
  - 68% of the probability is within 1 standard deviation.
  - 95% of the probability is within 2 standard deviation.
  - 99.7% of the probability is within 3 standard deviation.

- Example: For standard Normal



Calculating Normal Probabilities

- The Normal pdf is not directly integrable and we must use numerical integration. (Not practical to do routinely)
- Table is for the standard Normal. So, either standardize then look up or think about the problem in terms of standard deviations from the mean:

Normal Table

Table 4 Normal Curve Areas
Standard normal probability in right-hand tail
(for negative values of z, areas are found by symmetry)

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

From R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

Example

Example 17

A protein naturally produced in a rare tropical fruit can convert a sour taste into a sweet taste. Consequently, miraculin has the potential to be an alternative low-calorie sweetener. In Plant Science, a group of Japanese environmental scientists investigated the ability of a hybrid tomato plant to produce miraculin. For a particular generation of the tomato plant, the amount of miraculin produced (Y, measured in micrograms per gram of fresh weight) had a mean of 105.3 and a standard deviation of 8.0. Assume that Y is normally distributed.

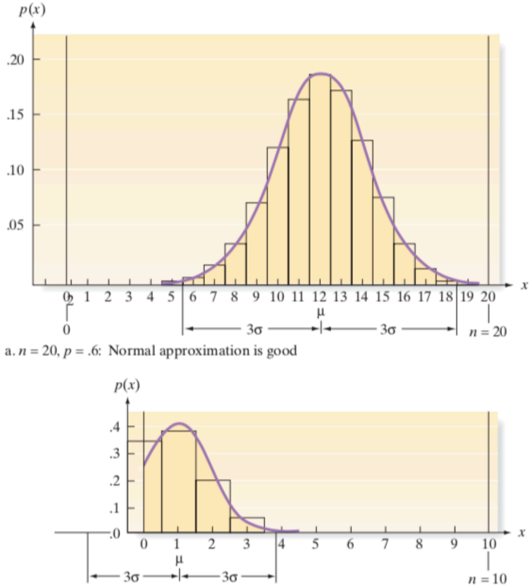
Example

- Find $P(Y > 120)$.
- Solution: $P(Y > 120) = P(Z > \frac{120-105.3}{8}) = P(Z > 1.838) = 0.033$
- Find $P(100 < Y < 110)$.
- Solution:
  $P(100 < Y < 110) = P(\frac{100-105.3}{8} < Z < \frac{110-105.3}{8}) = P(-0.663 < Z < 0.587) = 0.468$
- Find the value a for which $P(Y < a) = 0.25$
- Solution: $P(Z < \frac{a-105.3}{8}) = 0.25$. Using the standard normal Table, normal probability 0.25 corresponds to -0.674. Thus, $\frac{a-105.3}{8} = -0.674$ or $a = 99.904$

Approximating a Binomial Distribution with a Normal Distribution

- When the discrete binomial random variable can assume a large number of values, the calculation of its probabilities may become tedious.
- When n is large, a normal probability distribution may be used to provide a good approximation to the probability distribution of a binomial random variable.
- The normal distribution will not always provide a good approximation to binomial probabilities.
  A useful rule of thumb to determine when n is large enough: The interval $\mu \pm 3\sigma$ should lie within the range of the binomial random variable Y (i.e., from 0 to n)

Binomial Approximation



a. $n = 20$, $p = .6$: Normal approximation is good

b. $n = 10$, $p = .1$: Normal approximation is poor

Example (Exercise)

One problem with any product that is mass produced (e.g., a graphing calculator) is quality control. The process must be monitored or audited to be sure that the output of the process conforms to requirements. One monitoring method is lot acceptance sampling, in which items being produced are sampled at various stages of the production process and are carefully inspected. The lot of items from which the sample is drawn is then accepted or rejected on the basis of the number of defectives in the sample. Lots that are accepted may be sent forward for further processing or may be shipped to customers; lots that are rejected may be reworked or scrapped.

Example

Suppose that a manufacturer of calculators chooses 200 stamped circuits from the day's production and determines Y, the number of defective circuits in the sample. Suppose that up to a 6% rate of defectives is considered acceptable for the process.

- Find the mean and standard deviation of Y, assuming that the rate of defectives is 6%.
- Use the normal approximation to determine the probability that 20 or more defectives are observed in the sample of 200 circuits (i.e., find the approximate probability that $Y \geq 20$).

Gamma Distribution

- Note that, in addition to the normal distribution being symmetric about its mean, negative observations are (always) possible.
- But lots of real-world phenomenon can only be positive and come from skewed distributions.
- Examples:
  - Time between failures.
  – Distance to shell impact.
  - Time to receipt of repair part.

Defining the Gamma Distribution

Definition 18

A random variable Y has a Gamma distribution with parameters $\alpha > 0$ (shape parameter) and $\beta > 0$ (scale parameter) if and only if the density of Y is

$$f(y) = \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \qquad 0 \leq y < \infty$$

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

Theorem 19

If Y has the Gamma distribution, then $E(Y) = \alpha\beta$ and $V(Y) = \alpha\beta^2$.

Exponential Distribution

Definition 20

A random variable Y has an Exponential distribution with parameter $\beta > 0$ if and only if the density of Y is

$$f(y) = \frac{1}{\beta} e^{-y/\beta}, \qquad 0 \leq y < \infty$$

This is a Gamma distribution with parameters $(\alpha = 1, \beta)$.

Theorem 21

If Y has the Exponential distribution, then $E(Y) = \beta$ and $V(Y) = \beta^2$.

Notes

- If β is measured as the average number of time units per event then $1/\beta$ is the rate of events: the expected number of events per time period.
- If Y has the Exponential distribution, then the CDF of Y has a closed-form

$$F(y) = P(Y \leq y) = 1 - e^{-y/\beta}, \qquad y > 0$$

- Exponential distribution has the memoryless property: If Y has an Exponential distribution, with parameter $\beta > 0$, then

$$P(Y > a + b | Y > a) = P(Y > b)$$

Example: Hospital patient interarrival times

The length of time between arrivals at a hospital clinic has an approximately exponential probability distribution. Suppose the mean time between arrivals for patients at a clinic is 4 minutes.

- What is the probability that a particular interarrival time (the time between the arrival of two patients) is less than 1 minute?
- Solution: $P(Y \leq 1) = F(1) = 1 - e^{-1/4} = 0.221$
- What is the probability that the next four interarrival times are all less than 1 minute?
- What is the probability that an interarrival time will exceed 10 minutes?

Example: Hospital patient interarrival times

- What is the probability that the next four interarrival times are all less than 1 minute?
- Solution: Since the interarrival times are independent, we have

$$[P(Y \leq 1)]^4 = 0.221^4 = 0.00239$$

- What is the probability that an interarrival time will exceed 10 minutes?
- Solution: $P(Y > 10) = 1 - F(10) = 1 - [1 - e^{-10/4}] = e^{-10/4} = 0.0821$

Chi-Square Distribution

Definition 22

A random variable Y has a Chi-square distribution with ν degrees of freedom ($\mathcal{X}^2_\nu$) if and only if Y is a gamma distributed r.v. with parameters $\alpha = \nu/2$ and $\beta = 2$.

Theorem 23

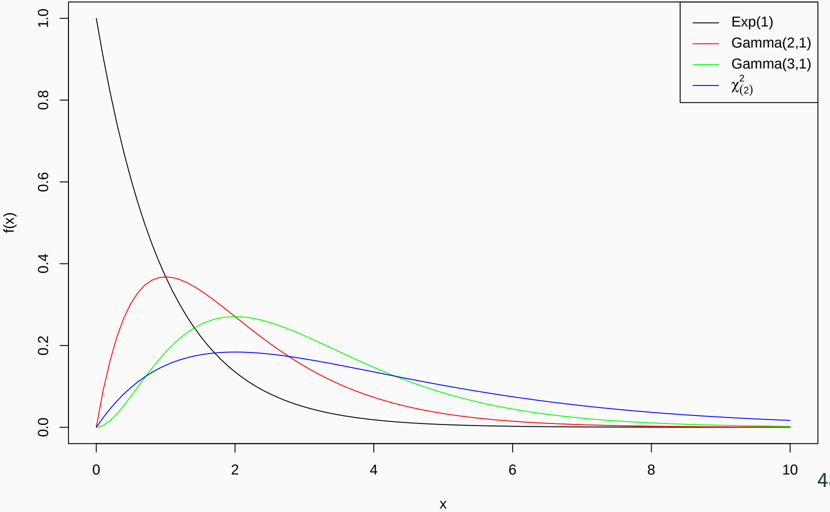
If Y has $\mathcal{X}^2_\nu$ distribution, then $E(Y) = \nu$ and $V(Y) = 2\nu$.

R code:

```
curve(dgamma(x, shape = 1, scale = 1), from = 0, to = 10, ylab = "f(x)")
curve(dgamma(x, shape = 2, scale = 1), add = TRUE, col = "red")
curve(dgamma(x, shape = 3, scale = 1), add = TRUE, col = "green")
curve(dgamma(x, shape = 4/2, scale = 2), add = TRUE, col = "blue")
legend("topright",legend = c("Exp(1)", "Gamma(2,1)", "Gamma(3,1)",expression(chi[(2)]^2)),
       col = c("black","red","green","blue"), lty = 1)
```

Gamma Distribution Curves

```
curve(dgamma(x, shape = 4/2, scale = 2), add = TRUE, col = "blue")
legend("topright",legend = c("Exp(1)", "Gamma(2,1)", "Gamma(3,1)",expression(chi
      col = c("black","red","green","blue"), lty = 1)
```



Beta Distribution

Definition 24

A random variable Y has a Beta distribution with parameters $\alpha > 0$ and $\beta > 0$ if and only if the density of Y is

$$f(y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}, \qquad 0 \leq y \leq 1$$

where

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1} dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Mean and Variance of Beta Distribution

Theorem 25

If Y has a Beta distribution, then $E(Y) = \frac{\alpha}{\alpha+\beta}$ and

$$V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

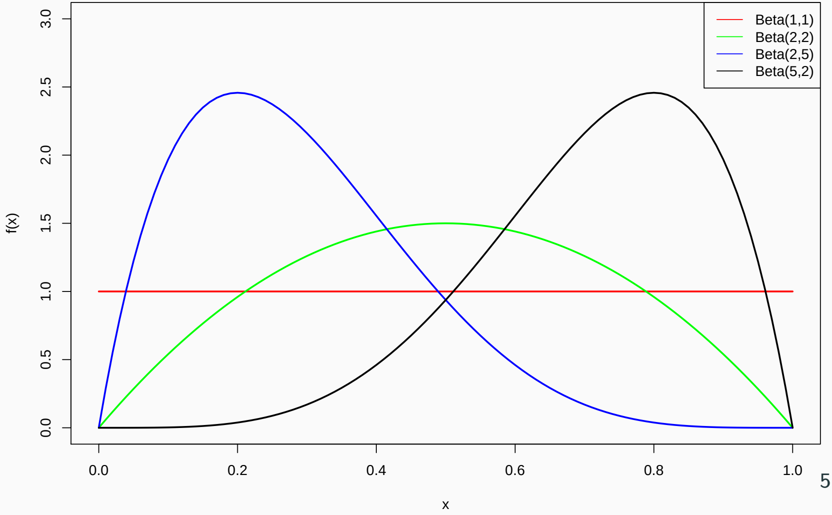
- Often used to model proportions (Example: The proportion of bacteria that die during a lab test).

R code

```
curve(dbeta(x, shape1 = 1, shape2 = 1), from = 0, to = 1,
      col = "red",lwd = 2, ylab = "f(x)",ylim = c(0, 3))
curve(dbeta(x, shape1 = 2, shape2 = 2), add = TRUE, col = "green",lwd = 2)
curve(dbeta(x, shape1 = 2, shape2 = 5), add = TRUE, col = "blue",lwd = 2)
curve(dbeta(x, shape1 = 5, shape2 = 2), add = TRUE, col = "black",lwd = 2)
legend("topright", legend = c("Beta(1,1)", "Beta(2,2)", "Beta(2,5)","Beta(5,2)"),
       col = c("red","green","blue","black"), lty = 1)
```

Beta Distribution Curves

```
curve(dbeta(x, shape1 = 5, shape2 = 2), add = TRUE, col = "black",lwd = 2)
legend("topright", legend = c("Beta(1,1)", "Beta(2,2)", "Beta(2,5)","Beta(5,2)"
       col = c("red","green","blue","black"), lty = 1)
```



Tchebysheff's Theorem

- Recall
$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$
- As with discrete random variables, Tchebysheff's Theorem provides a conservative lower bound for the probability that an observation falls in the interval $\mu \pm k\sigma$.
- It applies to any probability distribution, including continuous distributions.

Moments and Moment-Generating Functions

- The parameters μ and σ are meaningful numerical descriptive measures that locate the center and describe the spread associated with the values of a random variable.
- We now consider a set of numerical descriptive measures that (at least under certain conditions) uniquely determine $p(y)$.
- A major use of moments is to approximate the probability distribution of a random variable (usually an estimator or a decision maker).

Definiting kth Moment of a Random Variable

Definition 26

The kth moment of a random variable Y for $k = 1, 2, 3, ...$ taken about the origin is defined to be $E(Y^k)$ and is denoted by μ'_k.

Definition 27

The kth moment of a random variable Y taken about its mean, or the kth central moment of Y, is defined to be $E[(Y - \mu)^k]$ and is denoted by μ_k.

Moment-Generating Functions

Definition 28

The moment-generating function $m(t)$ for a random variable Y is defined to be $m(t) = E(e^{tY})$. We say that a moment-generating function for Y exists if there exists a positive constant b such that $m(t)$ is finite for $|t| \leq b$.

Application

The moment-generating function possesses two important applications.

- First, if we can find $E(e^{tY})$, we can find any of the moments for Y.
- The second (but primary) application of a moment-generating function is to prove that a random variable possesses a particular probability distribution $p(y)$. If $m(t)$ exists for a probability distribution $p(y)$, it is unique.

Application

Theorem 29

If $m(t)$ exists, then for any positive integer k

$$\left.\frac{d^k m(t)}{dt^k}\right|_{t=0} = m^{(k)}(0) = \mu_k'$$

In other words, if you find the kth derivative of $m(t)$ with respect to t and then set $t = 0$, the result will be μ_k' .

Example

Let $m(t) = (1/6)e^t + (2/6)e^{2t} + (3/6)e^{3t}$. Find the following:

- $E(Y) =$? Solution: $E(Y) = \frac{dm(t)}{dt}|_{t=0} = 14/6$
- $V(Y) =$? Solution: First,
  $E(Y^2) = \frac{d^2m(t)}{dt}|_{t=0} = 36/6 = 6$. Hence,
  $V(Y) = 6 - (14/6)^2 = 0.556$
- The distribution of Y
- Solution: From the definition of $m(t)$, i.e $m(t) = E(e^{tY})$, it is easy to see that

y	1	2	3
p(y)	1/6	2/6	3/6

What We Have Learned:

- Learned about continuous random variables, probability distributions, expected value, variance and standard deviation.
- Defined specific distributions
  - Uniform
  - Normal
  - Gamma Family (Exponential, Chi-square)
  - Beta
- Defined and applied Tchebycheff's Theorem
- Moments and moment Generating Functions