Random Variables and Distributions

Brian Vegetabile

2016 Statistics Bootcamp
Department of Statistics
University of California, Irvine

September 7th, 2016

A Few Definitions for Random Variables

Definition (Random Variable [Casella and Berger, 2002])

A random variable is a function from a sample space into the real numbers.

A Few Definitions for Random Variables

Definition (Random Variable [Schervish, 2014])

Consider an experiment for which the sample space is denoted by S. A real-valued function that is defined on the space S is called a random variable. In other words, in a particular experiment a random variable X would be some function that assigns a real number $X(s)$ to each possible outcome $s \in S$.

A Few Definitions for Random Variables

Definition (Random Variable [Dudewicz and Mishra, 1988])

A random variable X is a real-valued function with domain Ω [i.e. for each $\omega \in \Omega, X(\omega) \in \mathcal{R} = \{y : -\infty < y < +\infty\}$]. An n-dimensional random variable (or n-dimensional random vector, or vector random variable), $\mathbf{X} = (X_1, \ldots, X_n)$ is a function with domain Ω and range in Euclidean n-space $\mathcal{R}^n$. [i.e. for each $\omega \in \Omega, \mathbf{X}(\omega) \in \mathcal{R}^n = \{(y_1, \ldots, y_n) : -\infty < y_i < +\infty, (1 \leq i \leq n)\}$]

Example - Random Variables

- We can consider a random variable X which is the sum obtained from rolling two fair dice.
- Each die is capable of taking on the values $Roll_i = \{1, 2, 3, 4, 5, 6\}$, thus the sample space Ω is

$$\Omega = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

- And thus we can map values from the sample space into the reals. For example,

$$X((1,6)) = 7$$

Idea of "Induced Probabilities" I

- We now show how probabilities can be created for random variables, the following is from [Casella and Berger, 2002]
- This definition will be investigated more in higher level probability and statistics courses

Idea of “Induced Probabilities” II

- From Casella and Berger, consider the sample space Ω, with probability function P and we define a random variable X with range $\mathcal{X}$.
- We can define a probability function P_X on $\mathcal{X}$ in the following way.
- We will observe $X = x_i, x_i \in \mathcal{X}$ if and only if the outcome of the random experiment is an $\omega_j \in \Omega$, such that $X(\omega_j) = x_i$.
- Thus,

$$P_X(X = x_i) = P(\{\omega_j \in \Omega : X(\omega_j) = x_i\})$$

Example - 'Induced' Probabilities I

- We can consider a random variable X which is the sum obtained from rolling two fair dice.
- The sample space for the experiment is $\Omega = \{(i,j) : i,j = 1,2,3,4,5,6\}$.
- We have a probability function P, which describes the chance that each outcome in Ω is possible. That is

$$P((i,j)) = \frac{1}{36} \text{ for any combination of } i,j$$

since we are considering fair dice.

Example - 'Induced' Probabilities II

- Consider the random variable X, such that for each $\omega \in \Omega$,
$$X(\omega) = X((i,j)) = i + j$$
Therefore X can take on the values $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.
- Now using this fact, we can create probabilities for the random variable X. For example, we can find the probability that $X = 4$.

$$\begin{aligned} P_X(X = 4) &= P(\{\omega_j \in \Omega : X(\omega_j) = 4\}) \\ &= P(\{(1,3), (2,2), (3,1)\}) \\ &= P((1,3)) + P((2,2)) + P((3,1)) \\ &= \frac{3}{36} = \frac{1}{12} \end{aligned}$$

Example - 'Induced' Probabilities III

- Therefore using the original probability function on the sample space we can find probabilities for the random variable X.
- P_X is an induced probability function on $\mathcal{X}$.

Discrete Random Variables

- Random variables which can only take on at most a countable number of possible values are considered to be discrete random variables.
- Each one of these possible values will be assigned some non-negative probability, therefore we can intuitively as placing 'mass' of probability at each possible value.

Discrete Random Variables

Definition (probability mass function)

If a random variable X is a discrete distribution (that is it takes on only a countable number of different values) then the probability mass function of X is defined as the function f such that for every real number x

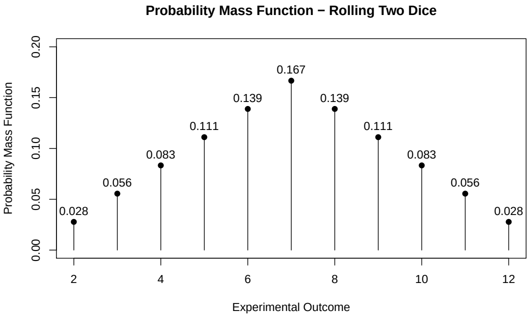
$$f(x) = P_X(X = x) \text{ for all } x$$

Since f is a probability function it follows that $f(x) \geq 0$ for all x and

$$\sum_{\text{all } x} f(x) = 1$$

Example - Probability Mass Function for Rolling Two Dice

- Again consider the random variable X which is the sum obtained from rolling two fair dice.
- Clearly the random variable takes on a countable number of different values
- Thus we can construct the probability mass function for this random variable.



Continuous Random Variables I

- Continuous random variables are concerned with probability on intervals.
- From Degroot/Schervisch, a random variable X has a continuous distribution, or is a continuos random variable, if there exists a non-negative function f, defined on the real line, such that for every subset A of the real line, the probability that X takes a value in A is the integral over the set A.

$$P_X(X \in A) = \int_{x \in A} f(x)dx$$

Continuous Random Variables II

- When dealing with intervals $(a, b]$ this becomes

$$P_X(a < X \leq b) = \int_a^b f(x)dx$$

- The function f is called the probability density function of the random variable X. Again $f(x) \geq 0$ for all x and

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- Additionally recall from calculus that $\int_a^a f(x)dx = 0$ for any x. This further implies that

$$P_X(X = x) = 0$$

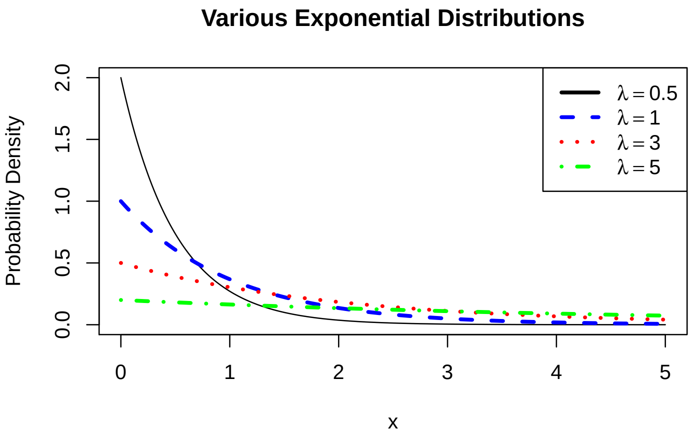
for a continuous random variable.

Example - Continuous Random Variable I

- Consider a random variable that measures the time between events occurring in an interval of time.
  - Notice this is a measure of time and therefore we will only expect probability density on the non-negative reals.
- Such a random variable is called an exponential random variable and has the following density

$$f(x|\lambda) = \begin{cases} \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Example - Continuous Random Variable II



A Quick Aside: Observable Data vs. Unobservable Parameters

- Before continuing this is a good time to point out a fundamental issue in statistics
- In our dice example, we are able to make assumptions, based on geometry or other things, which allow us to *apriori* know the probability distribution for each experimental outcome.
- In the continuous distribution example, notice the parameter λ in the distribution.
- It is often impossible to know these parameters beforehand and thus our goal eventually will be to estimate these parameters.
- Often we have a set of "observed data" and we want to help us learn about the "unobservable parameters"

Cumulative Distribution Functions I

- Once we understand the probability functions for random variables we can begin to talk about other functions of, and properties of, the random variables.
- The first major function that is considered for every random variable is its cumulative distribution. This is one of the first places that integration will come into play.

Cumulative Distribution Functions II

Definition (Cumulative Distribution Function)

The distribution function F_X of a random variable X is a function defined for each real number x as follows:

$$F(x) = P_X(X \leq x) \text{ for all } x$$

- Additionally based on our previous definitions of the density function this is equivalent to

$$F(x) = P_X(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Relationship between Mass/Density Functions And Distribution Functions

- From the Fundamental Theorem of Calculus if the random variable is continuous we have that

$$F'(x) = \frac{d}{dx}F(x) = f(x)$$

Properties of the CDF

i) The function $F_X(x)$ is nondecreasing as x increases; that is, if $x_1 < x_2$, then $F_X(x_1) \leq F_X(x_2)$.

ii) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$

iii) $F_X(x)$ is right continuous.

Example - Discrete Random Variable CDF I

- Consider the function

$$f(x) = \begin{cases} 0.1 & \text{if } x = 0 \\ 0.2 & \text{if } x = 1 \\ 0.7 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

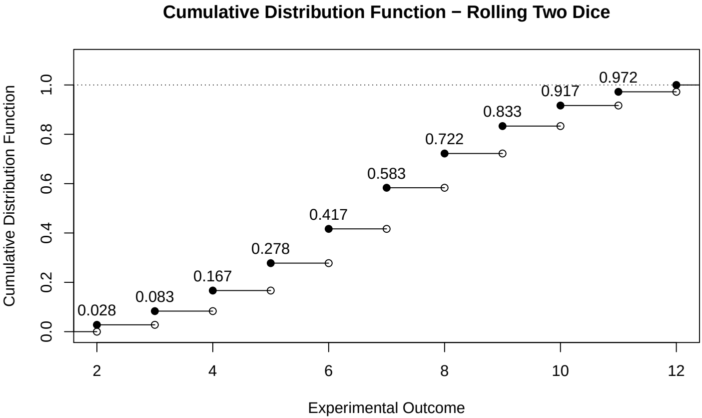
- Clearly this is a probability mass function since $f(x) \geq 0$ and $\sum_{i=0}^{2} f(x) = 1$
- To find the distribution function must have a right continuous function where we find $P(X \leq x)$ for all x

Example - Discrete Random Variable CDF II

- Therefore, by the mass function we have that $P(X < 0) = 0$, and we see that $P(X \leq 0) = 0.1$.
- Now since the next 'mass jump' is at $x = 1$ the function remains constant up until that point and for example $P(X \leq 0.5) = 0.1$
- Once $x = 1$ though we add mass and have $P(X \leq 1) = 0.3$ and finally $P(X \leq 2) = 1$.
- We can see how this is a 'cumulative' function.

Example - Discrete Random Variable CDF

- Continuing with our rolling two dice example we can create the CDF of this distribution by considered $P(X \leq x)$ for all x.
- Therefore we obtain the following function,



Example - Continuous Random Variable CDF I

- For the exponential distribution we can integrate the probability density function to obtain the cumulative distribution function
- Recall $f(x|\lambda) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right)$ and $F(x) = \int_{-\infty}^{x} f(x)dx$.

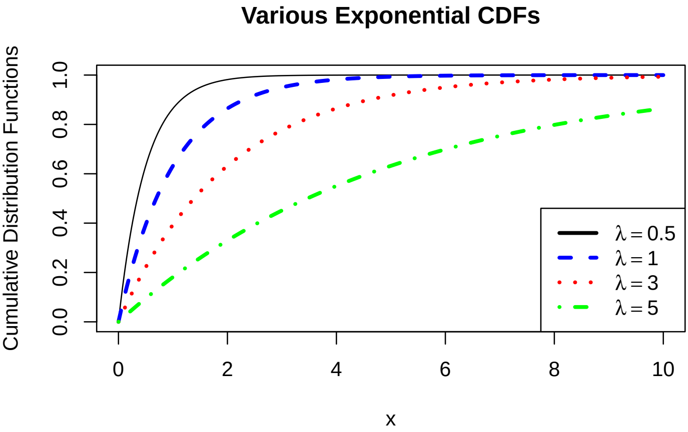
Example - Continuous Random Variable CDF II

- Thus

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(x)dx) = \int_{0}^{x} \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) dx \\
&= -\exp\left(-\frac{x}{\lambda}\right)\Big|_{0}^{x} = -\exp\left(-\frac{x}{\lambda}\right) + 1 \\
&= 1 - \exp\left(-\frac{x}{\lambda}\right)
\end{aligned}
$$

- We can use the CDF is this case to find the probability that the time between events is less than some value (or greater than).
- This can inform us about rare event times.

Example - Continuous Random Variable CDF III



Expectations of Random Variables

- The distribution of a random variable X contains all of the probabilistic information about X.
- Often we summarize the random variable by certain measures, namely expectations, variances, means, modes, etc.
- We introduce the first such summary, the expectation of the random variable

Defining the Expected Value

- This value is sometimes referred to as the *mean* of the random variable.

Definition (Expected Value)

The expected value of a random X, denoted $E(X)$ is defined

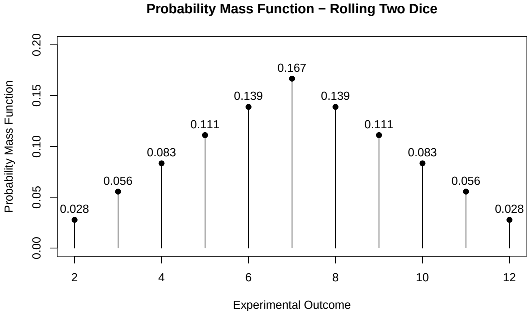
$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

For a discrete random variable this would become

$$E(X) = \sum_{\text{all } x} x f(x)$$

Example - Discrete Random Variable I

- We continue investigating our created random variables.
- As an intuition the expected value can be thought of as a “balancing point” of the distribution.



Example - Discrete Random Variable II

- Performing our calculation

$$
\begin{aligned}
E(X) &= \sum_x x f(x) = 2 \times 0.028 + 3 \times 0.056 + \ldots \\
&= 7
\end{aligned}
$$

- One thing worth noting is that It is NOT where the most mass is, as we will see in the continuous example.

Example Expected Value of a Continuous Random Variable I

- For an exponential random variable, this can be thought of as the expected time between two events occurring.
- Calculating...

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{\infty} \frac{x}{\lambda} \exp\left(-\frac{x}{\lambda}\right)$$

Example Expected Value of a Continuous Random Variable II

- We can do this by integration by parts, where '$\int u dv = uv - \int v du'$ with the following

$$
\begin{aligned}
u &= \frac{x}{\lambda} & dv &= \exp\left(-\frac{x}{\lambda}\right) \\
du &= \frac{1}{\lambda} & v &= -\lambda \exp\left(-\frac{x}{\lambda}\right)
\end{aligned}
$$

Example Expected Value of a Continuous Random Variable III

- Thus,

$$
\begin{aligned}
E(X) &= \int_0^\infty \frac{x}{\lambda} \exp\left(-\frac{x}{\lambda}\right) \\
&= -x \exp\left(-\frac{x}{\lambda}\right)\Big|_0^\infty + \int_0^\infty \exp\left(-\frac{x}{\lambda}\right) dx \\
&= -\lambda \exp\left(-\frac{x}{\lambda}\right)\Big|_0^\infty \\
&= \lambda
\end{aligned}
$$

- As we saw from our earlier example, the maximum of the distribution occurs at 0, which does not correspond to the expected value of this distribution.

An Important Property of Expectations I

- We now present a useful relationship for expectations
- Claim: $Y = aX + b$, then $E(Y) = aE(X) + b$.

An Important Property of Expectations II

- Claim: $Y = aX + b$, then $E(Y) = aE(X) + b$.
- Proof:

$$\begin{aligned} E(Y) &= E(aX+b) = \int (ax+b) f(x) dx \\ &= a \int x f(x) dx + b \int f(x) dx \\ &= aE(X) + b \end{aligned}$$

Generalizing Expectations of Random Variables

- We are not restricted to only finding the expected value for the random variable X, we can also find the expected value of *functions* of the random variable X, that is some $g(X)$.

Definition (Law of The Unconcious Statistician)

The expected value of a function of the random value X, $g(X)$, denoted $E(g(X))$ is defined

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

For a discrete random variable this would become

$$E(g(X)) = \sum_{\text{all } x} g(x)f(x)$$

Variance of Random Variables I

- One function that is ultimately of interest is the variance of a random variable
- The variance allows us to quantify the variability or the spread of a distribution.

Definition (Variance)

The variance of a random variable is defined as follows

$$Var(X) = E[(X - E(X))^2]$$

Often the expected value is defined as $E(X) = \mu$ and therefore this is often written in textbooks as

$$Var(X) = E[(X - \mu)^2]$$

Variance of Random Variables II

The equation for variance can be simplified to the following useful relationship.

- Claim : $Var(X) = E(X^2) - (E(X))^2$

Variance of Random Variables III

- Proof:

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= \int (x^2 - 2\mu x + \mu^2) f(x) dx \\
&= \int x^2 f(x) dx - \int 2\mu x f(x) dx + \int \mu^2 f(x) dx) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

A Simple Property of Variance I

- Claim: $Var(aX + b) = a^2 Var(X)$

A Simple Property of Variance II

- Claim: $Var(aX+b) = a^2Var(X)$
- Proof: We know $E(aX+b) = aE(X)+b$, thus

$$\begin{aligned} Var(aX+b) &= E((aX+b-E(aX+b))^2) \\ &= E((aX+b-aE(X)+b)^2) \\ &= E(a^2(X-E(X))^2 \\ &= a^2E((X-E(X))^2) = a^2Var(X) \end{aligned}$$

Moments of Random Variables

- Related to both the mean and variance of a random variable are it's moments.
- Moments will become useful for simple estimation strategies

Definition (Moment of a Random Variable)

For each random variable X and every positive integer k, $E(X^k)$ is called the k^{th} moment of X.

- We can also talk about **Central Moments**

Definition (Central Moment of a Random Variable)

Suppose that X is a RV such that $E(X) = \mu$, then for every positive integer k, $E((X-\mu)^k)$ is called the k^{th} central moment of X.

Moment Generating Functions

- Finally, one of the most important properties of a random variable is its moment generating function.
- If two random variables have moment generating functions that exist and are equal, then they have the same distribution and similarly the converse is true.
- MGFs will be useful for showing convergence of random variables, they're used tor finding the distribution of a sum of independent and identically distributed random variables, and they play important role in the Central Limit Theorem.

Moment Generating Functions

Definition (Moment Generating Functions)

The moment-generating function of a random variable X is defined for every real number t by $M_X(t) = E(e^{tX})$.

- MGF's can be used to 'generate' the moments of the random variable in the following way:

$$EX^n = M_{X^{(n)}}(0) \equiv \frac{d^n}{dt^n} M_X(t)\bigg|_{t=0}$$

MGF for the Exponential Distribution I

- Want to find $E(e^{tX})$

$$E(e^{tX}) = \int_0^\infty \exp(tx) \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) dx$$

MGF for the Exponential Distribution II

$$
\begin{aligned}
E(e^{tX}) &= \int_0^\infty \exp(tx)\frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right)dx \\
&= \int_0^\infty \frac{1}{\lambda}\exp\left(tx-\frac{x}{\lambda}\right)dx \\
&= \int_0^\infty \frac{1}{\lambda}\exp\left(-x\frac{1-\lambda t}{\lambda}\right)dx \\
&= -\frac{1}{\lambda}\frac{\lambda}{1-\lambda t}\exp\left(-x\frac{1-\lambda t}{\lambda}\right)\Big|_0^\infty \\
&= 0+\frac{1}{1-\lambda t} \\
&= \frac{1}{1-\lambda t}
\end{aligned}
$$

Extending to Multiple Random Variables I

- Clearly, we will want to extend these concepts to multiple random variables.
- For example, we may like to consider the joint distribution for two random variables X and Y.
- Therefore we require a function f such that such that for every subset A of the sample space, the probability that X and Y, the pair (x, y), takes a value in A, is the integral over the set A.

$$P(X, Y \in A) = \iint_A f(x,y)dxdy$$

Extending to Multiple Random Variables II

- When we are considering two random variables, this is considered a bivariate distribution.
- When we consider more than two random variables this is considered a multivariate distribution.

$$
\begin{aligned}
& P(X_1, X_2, \dots X_n \in A) \\
= & \int \dots \iint_A f(x_1, x_2, \dots x_n) dx_1 dx_2 \dots dx_n
\end{aligned}
$$

- Multivariate random variables have corresponding distribution functions, expectations, moments, etc.

Marginal Distributions

- We now consider properties of joint distribution functions, which we will illustrate mainly using bivariate distributions.
- First we consider finding the marginal distribution of random variables.
- Consider two random variables X and Y and their joint distribution $f(x, y)$.
- The marginal distribution of X and Y are given as follows

$$f_X(x) = \int_{y \in \mathcal{R}} f(x, y) dy \qquad f_Y(y) = \int_{x \in \mathcal{R}} f(x, y) dx$$

Example - Joint Distribution I

- Consider the density

$$f(x,y) = x^2(1+y) + \frac{3}{2}y^2 \text{ for } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1$$

- We first ask is this a proper density?
  - Need $f(x,y) \geq 0$ for any $(x,y) \in \Omega = [0,1] \times [0,1]$
  - Need $\iint f(x,y)dxdy = 1$

Example - Joint Distribution II

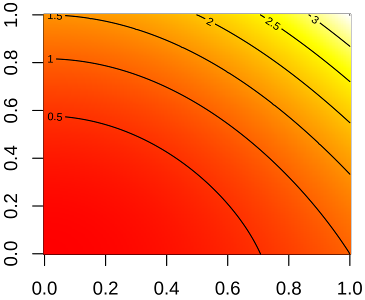
- Taking derivatives

$$\left(\frac{d}{dx}f(x,y), \frac{d}{dy}f(x,y)\right) = \left(2x(1+y), x^2 + 3y\right)$$

We see that the extreme values can be found at $(0,0)$ which occurs at a corner.

Example - Joint Distribution III

- Additionally, plotting we see that the distribution is strictly increasing



Example - Joint Distribution IV

- Now interested in $\iint f(x,y)dxdy = 1$

$$
\begin{aligned}
\int_0^1 \int_0^1 x^2(1+y) + \frac{3}{2}y^2 dydx &= \int_0^1 x^2 \left(y + \frac{y^2}{2}\right) + \frac{3}{2}\frac{y^3}{3}\Big|_0^1 dx \\
&= \int_0^1 \frac{3}{2}x^2 + \frac{1}{2}dx \\
&= \frac{3}{2}\frac{x^3}{3} + \frac{1}{2}x\Big|_0^1 \\
&= \frac{1}{2} + \frac{1}{2} = 1
\end{aligned}
$$

Example - Joint Distribution V

- Now let us find the marginal distributions

$$
\begin{aligned}
f(y) &= \int_0^1 x^2(1+y) + \frac{3}{2}y^2 dx \\
&= \left.\frac{x^3}{3}(1+y) + \frac{3}{2}y^2 x\right|_0^1 \\
&= \frac{1}{3}(1+y) + \frac{3}{2}y^2 \\
f(x) &= \int_0^1 x^2(1+y) + \frac{3}{2}y^2 dy \\
&= \left.x^2\left(y + \frac{y^2}{2}\right) + \frac{y^3}{2}\right|_0^1 \\
&= \frac{3}{2}x^2 + \frac{1}{2}
\end{aligned}
$$

Conditioning on Random Variables

- Once we have a way to calculate the marginal distribution of a random variable, we can begin to think of conditioning on random variables.
- This is similar to conditioning in probability that we learned about.

Definition (Conditional Density Function)

Consider two random variables X and Y, we define the conditional density of X given $Y = y$ denoted $X|Y = y$ to be

$$f_{X|Y=y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

- Conditional distributions will become increasingly important in Bayesian statistics

Independence between Random Variables

- Consider random variables $X_1, X_2, \ldots, X_n$, we say the random variables are independent if

$$f(x_1, x_2, \ldots x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n)$$

- That is random variables are independent if we can factorize the joint distribution into a product of the individual marginal distributions.
- If each $f_{X_i}(\cdot)$ is the same, we say the they are identically distributed.
- Independence and identical distributions will come into use for describing random samples

Expectations of Joint Distributions

- Similar to the univariate case, we can discuss expectations in terms of the joint distribution.
- For a bivariate distribution we can obtain the following relationships

$$
\begin{aligned}
E(X) &= \int\int xf(x,y)dxdy = \int xf(x)dx \\
E(Y) &= \int\int yf(x,y)dxdy = \int yf(y)dy \\
E(XY) &= \int\int xyf(x,y)dxdy
\end{aligned}
$$

- Additionally, if X and Y are independent we can see that $E(XY) = E(X)E(Y)$

Covariance & Correlation between Random Variables

- Summaries of the random variables allow us to understand how they variables behave (expectations), but additionally we would like to know how their behavior is related
  - This is especially true if the random variables are not independent.
- Both the covariance and correlation between random variables attempts to measure the dependence between random variables

Defining Covariance between Random Variables I

- Covariance allows us to begin to talk about the tendency of two variables to vary together rather than independently.
- If greater values of one random variable correspond to greater values of the other random variable (similar for smaller values), then the covariance will be positive.
- If greater values of one random variable correspond to lesser values of another random variable, then the covariance will be negative.
- The magnitude of covariance is often uninterpretable

Defining Covariance between Random Variables II

Definition

Covariance Consider the random variables X and Y and let $E(X) = \mu_X$ and $E(Y) = \mu_Y$ we define the covariance between X and Y to be

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

- Similar to the univariate case, it can be shown that the above reduces to

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

which is much easier to work with in application.

Correlation between Random Variables I

- As stated in the previous slides, the magnitude of covariance is not helpful in understanding how to random variables values are related.
- This is due to the fact that covariance is sensitive to the relative magnitudes of X and Y.
- To remedy this, the notion of correlation is not driven by arbitrary changes in the scales of the random variables.

Correlation between Random Variables II

Definition (Correlation)

Consider the random variables X and Y, with finite positive variances $Var(X)$ and $Var(Y)$, then the correlation between X and Y is given as follows

$$\rho(X,Y) = \frac{Cov(X,Y)}{Var(X)Var(Y)}$$

- It can be shown that the correlation is restricted between -1 and 1 thus

$$-1 \leq \rho(X,Y) \leq 1$$

References

George Casella and Roger L Berger. *Statistical inference*. Second edition, 2002.

EJ Dudewicz and SN Mishra. Modern mathematical statistics. john wilsey & sons. *Inc., West Sussex*, 1988.

Mark J Schervish. *Probability and Statistics*. 2014.