CSE 312: Foundations of Computing II
Advanced Topics Session #4

The Gamma Distribution, MGF's, Order Statistics, Conditional Expectation

Lecturer: Alex Tsun

Date: May 9, 2017

1 The Gamma Function

We define the **gamma function**,

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$$

Exercise: Show that $\Gamma(r) = (r-1)\Gamma(r-1)$. (Hint: Use integration by parts).
Solution: We use integration by parts with the following substitution to get

$$u = x^{r-1}, du = (r-1)x^{r-2}dx, v = -e^{-x}, dv = e^{-x}dx$$
$$\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx = -[x^{r-1}e^{-x}]_0^\infty + (r-1)\int_0^\infty x^{r-2}e^{-x}dx = (r-1)\Gamma(r-1)$$

Exercise: Find $\Gamma(1)$.
Solution:

$$\Gamma(1) = \int_0^\infty e^{-x}dx = 1$$

Exercise: Let $n \in \mathbb{N}$ be a positive integer. Use the previous two exercises to conclude that $\Gamma(n) = (n-1)!$.
Solution: We use induction. The base case was the previous exercise. Suppose $\Gamma(n-1) = (n-2)!$ for some $n \in \mathbb{N}$. Then, by the first exercise and the induction hypothesis, $\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)! = (n-1)!$. So the gamma function attempts to fit a curve to extend factorial to all real numbers!

2 The Gamma Distribution

The **Poisson Process with parameter λ** is a stochastic process over an (uncountably) infinite interval of time, namely $[0, \infty)$ such that successes/events occur at an average rate of λ per unit time. The **Poisson distribution with parameter λ** measures the number of successes in a unit of time, and we say $X \sim Poi(\lambda)$. Recall that the Poisson distribution stems from the Binomial, when we let $n \to \infty, p \to 0$ in such a way that $np \to \lambda$. Now, we describe the **Exponential distribution with parameter λ** as the waiting time until the first event. Notice that if $\boldsymbol{X \sim Exp(\lambda)}$, then $\Omega_X = [0, \infty)$, and $F_X(x) = 1 - e^{-\lambda x}$.

Why is this? Let Y_x be the number of successes in the first x units of time of the Poisson process with parameter λ. Since λ is the rate per single unit of time, λx is the rate parameter for Y_x. So $Y_x \sim Poi(\lambda x)$.

$$P(X > x) = P(no\ events\ in\ the\ first\ x\ units\ of\ time) = P(Y_x = 0) = e^{-\lambda x}\frac{(\lambda x)^0}{0!} = e^{-\lambda x}$$
$$F_X(x) = P(X \leq x) = 1 - P(X > x) = 1 - e^{-\lambda x}$$

We can differentiate the cdf to get its pdf as $f_X(x) = \lambda e^{-\lambda x}$. Just like the geometric distribution, the exponential distribution represents the waiting time until the first success, so it is the continuous analog of the geometric distribution. More similarities: both are memoryless, that is $P(X > s + t | X > t) = P(X > s)$, so if you've waited at least t units of time, the probability you wait s more is the same as the probability you wait at least s units from the beginning. Furthermore, the expectation of $Geo(p)$ is $\frac{1}{p}$ and the expectation of $Exp(\lambda)$ is $\frac{1}{\lambda}$.

So what is the continuous analog of the negative binomial distribution? If $X_1, \dots, X_r$ are iid $Exp(\lambda)$, their sum $X = X_1 + \cdots + X_r$ is the waiting time until the r^{th} event, and we say that X has the **gamma distribution with parameters r and λ**, that is, $\boldsymbol{X \sim Gam(r, \lambda)}$. We call r the **shape parameter** and λ the **rate parameter**.

Exercise: Find the expectation of X and variance of X, if $X \sim Gam(r, \lambda)$.
Solution: Recall for $W \sim Exp(\lambda)$, we have $E[W] = \frac{1}{\lambda}$ and $Var(W) = \frac{1}{\lambda^2}$. Since $X = X_1 + \cdots + X_r$, where $X_i \sim Exp(\lambda)$, we use linearity of expectation to get

$$E[X] = E\left[\sum_{i=1}^{r} X_i\right] = \sum_{i=1}^{r} E[X_i] = \sum_{i=1}^{r} \frac{1}{\lambda} = \frac{r}{\lambda}$$

Since we have independence, we can use linearity of variance to get

$$Var(X) = Var\left(\sum_{i=1}^{r} X_i\right) = \sum_{i=1}^{r} Var(X_i) = \sum_{i=1}^{r} \frac{1}{\lambda^2} = \frac{r}{\lambda^2}$$

The probability density function of X is given by

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, x > 0$$

But we will not attempt to derive or explain the density. The gamma function is used to ensure the density integrates to 1. We will see this distribution later when we talk about parameter estimation, but for now, it is still useful to know about!

3 Moment Generating Functions

Let X be a random variable. We define the $\boldsymbol{k^{th}}$ **moment of** $\boldsymbol{X}$ **about** $\boldsymbol{c}$ as $E[(X-c)^k]$. If we don't say "about c", we assume $c = 0$. So for example, the first moment is $E[X] = \mu$ and the second moment about μ is the variance: $E[(X-\mu)^2]$. We define the **moment generating function (mgf/MGF) of** $\boldsymbol{X}$, as $M_X(t) = E[e^{tX}]$. It allows us to find moments easily, and uniquely defines a distribution.

If X is discrete,

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} p_X(x)$$

If X is continuous,

$$M_X(t) = E[e^{tX}] = \int_{\infty}^{\infty} e^{tx} f_X(x) dx$$

Here are some properties:

1. $M_X'(0) = E[X]$, $M_X''(0) = E[X^2]$, and in general, $M_X^{(n)}(0) = E[X^n]$, where $M_X^{(n)}(t)$ is the n^{th} derivative of $M_X(t)$.
2. If $X \perp Y$, then $M_{X+Y}(t) = M_X(t) M_Y(t)$.
3. For $a, b \in \mathbb{R}$, $M_{aX+b}(t) = e^{bt} M_X(at)$.
4. (Uniqueness) $f_X(s) = f_Y(s)$ for all $s \in \mathbb{R}$ if and only if $M_X(t) = M_Y(t)$ for all t in some ϵ-neighborhood of 0. (i.e. the MGF uniquely determines a distribution, just like the pdf/pmf and cdf).

Exercise: Prove property 1, assuming X is discrete. (The same proof applies for continuous rv's).
Solution:

$$M_X'(t) = \frac{d}{dt} \sum_x e^{tx} p_X(x) = \sum_x \frac{d}{dt}(e^{tx}) p_X(x) = \sum_x x e^{tx} p_X(x)$$

$$M_X'(0) = \sum_x x p_X(x) = E[X]$$

$$M_X''(t) = \frac{d^2}{dt^2} \sum_x e^{tx} p_X(x) = \sum_x \frac{d^2}{dt^2}(e^{tx}) p_X(x) = \sum_x x^2 e^{tx} p_X(x)$$

$$M_X''(0) = \sum_x x^2 p_X(x) = E[X^2]$$

This can easily be extended by induction to $M_X^{(n)}(0) = E[X^n]$.

Exercise: Prove properties 2 and 3, assuming X, Y are discrete. (The same proof applies for continuous rv's). You may use the fact that, if X and Y are independent, so are $g(X)$ and $h(Y)$ for any arbitrary functions g and h.

Solution:

$$M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$$

where we require independence for the third equality.

$$M_{aX+b}(t) = E\left[e^{t(aX+b)}\right] = E[e^{atX}e^{tb}] = e^{tb}E\left[e^{(at)X}\right] = e^{bt}M_X(at)$$

Exercise: Find the moment generating function of X if $X \sim Exp(\lambda)$, and use it to verify $E[X] = \frac{1}{\lambda}$.
Solution:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda\int_0^\infty e^{(t-\lambda)x}dx = \frac{\lambda}{t-\lambda}\left[e^{(t-\lambda)x}\right]_0^\infty = \frac{\lambda}{\lambda - t}$$

For the last equality, we needed to suppose $t < \lambda$ in order for the integral to converge, but it's okay because we only need M_X defined on a small neighborhood of 0.

$$M_X'(t) = \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right) = \frac{\lambda}{(\lambda - t)^2}$$
$$E[X] = M_X'(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

So moment generating functions are super useful and a nice way to find moments! It may not have been worth it in this case, but sometimes these are helpful. Again, also mgf's uniquely define a distribution as much as the pdf/pmf and cdf.

4 Order Statistics

Suppose $Y_1, \dots, Y_n$ are iid <u>continuous</u> random variables with common pdf f_Y and common cdf F_Y. We sort the Y_i's such that

$$Y_{min} \equiv Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)} \equiv Y_{max}$$

Notice we can't have equality because with continuous random variables, the probability that any two are equal is identically 0. Notice that each $Y_{(i)}$ is random variable as well! We call $Y_{(i)}$ the $\boldsymbol{i^{th}}$ **order statistic**: the i^{th} smallest in a sample of size n. So we are interested in finding the distribution of each order statistic, and properties such as expectation and variance as well.

We start with an example to find the distribution of Y_{max}, the largest order statistic. We start with the cdf:

$$F_{Y_{max}}(y) = P(Y_{max} \le y) = P\left(\bigcap_{i=1}^n Y_i \le y\right) = \prod_{i=1}^n P(Y_i \le y) = \prod_{i=1}^n F_Y(y) = F_Y^n(y)$$

where the first equality is by definition, the second is because the maximum is $\leq y$ if and only if all Y_i's are $\leq y$, the third is because of independence, and the rest are definitions.

Then, we can differentiate the cdf to find the density function:

$$f_{Y_{max}}(y) = F'_{Y_{max}}(y) = \frac{d}{dy}\left(F_y^n(y)\right) = nF_Y^{n-1}(y)f_Y(y)$$

We use a very informal argument to find the density of a general $Y_{(i)}$, $f_{Y_{(i)}}(y)$. One of the values needs to be exactly y, $i-1$ need to be smaller than y (with probability $F_Y(y)$), and the other $n-i$ need to be greater than y (with probability $1-F_Y(y)$). Notice also that there are 3 distinct types of objects: 1 that is exactly y, $i-1$ which are less than y, and $n-i$ which are greater. So we multiply by the multinomial coefficient $\binom{n}{i-1,1,n-i}$ to get

$$f_{Y_{(i)}}(y) = \binom{n}{i-1,1,n-i}[F_Y(y)]^{i-1}[1-F_Y(y)]^{n-i}f_Y(y)$$

Let's verify the result we got earlier:

$$f_{Y_{max}}(y) = f_{Y_{(n)}}(y) = \binom{n}{n-1,1,0}[F_Y(y)]^{n-1}[1-F_Y(y)]^0 f_Y(y) = nF_Y^{n-1}(y)f_Y(y)$$

5 Jointly Distributed Continuous Random Variables

We define things very similarly for continuous random variables, when we extend to the multivariate case. Suppose X and Y are continuous random variables. We call the **joint probability density function (joint pdf) of X and Y** $f_{X,Y}(x,y)$. Notice that we use an f instead of p – we reserve f for density function and p for mass functions. We say that X and Y are **independent** if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x, y. The **marginal density** (similar to the marginal pmf) is defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

The **joint cumulative distribution function (joint cdf)** is given by

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(t,s)dsdt$$

Similarly to the univariate case, we have

$$\frac{\partial^2}{\partial x \partial y}F_{X,Y}(x,y) = f_{X,Y}(x,y)$$

We summarize the transition from jointly discrete random variables to jointly continuous.

Multivariate: Discrete to Continuous:

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt$
Normalization	$\sum_x \sum_y p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Expectation	$E[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$
Conditional PMF/PDF	$p_{X \mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y = y] = \sum_x x p_{X \mid Y}(x \mid y)$	$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y) dx$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

1. Suppose X and Y are continuous random variables with joint density

$$f_{X,Y}(x,y) = \begin{cases} cxy^2, & x > 0, y > 0, x + y < 1 \\ 0, & otherwise \end{cases}$$

a) Write an equation which we can solve to find the value of c.

$$\int_0^1 \int_0^{1-x} cxy^2 dy dx = 1$$

b) Write an expression which we can solve to find $P(Y \geq X)$.

$$P(Y \geq X) = \int_0^{1/2} \int_x^{1-x} cxy^2 dy dx$$

c) Write an expression to find the marginal pdf, $f_X(x)$. Specify its value for all $x \in \mathbb{R}$.

$$f_X(x) = \int_0^{1-x} cxy^2 dy, for\ 0 < x < 1$$

d) Write an expression to find the joint CDF, $F_{X,Y}(s,t)$. Specify its value for all $s, t \in \mathbb{R}$.

$$F_{X,Y}(s,t) = \int_0^t \int_0^s cxy^2 dx dy, where\ s,t > 0, s + t < 1$$

e) Are X and Y independent?

No, they aren't defined over a rectangle.

f) Find an expression for $E[\sin(X^Y)]$.

$$E[\sin X^Y] = \int_0^1 \int_0^{1-x} \sin(x^y)\, cxy^2 dydx$$

g) Suppose V, W, X, Y, Z are jointly continuous with pdf $f_{V,W,X,Y,Z}(v,w,x,y,z)$. Write an expression for the marginal joint density, $f_{V,X,Z}(v,x,z)$.

$$f_{V,X,Z}(v,x,z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{V,W,X,Y,Z}(v,w,x,y,z) dwdy$$

2. Let (X,Y) have joint range $\Omega_{X,Y} = \{(x,y): x^2 + y^2 \leq 1\}$, the unit disk in $\mathbb{R}^2$. Find $f_{X,Y}(x,y)$ if X and Y are uniformly distributed on $\Omega_{X,Y}$. Are X and Y independent? If so, prove it. If not, does there exist any $f_{X,Y}(x,y)$ on $\Omega_{X,Y}$ such that X and Y are independent?

$$f_{X,Y}(x,y) = \frac{1}{\pi}, x^2 + y^2 \leq 1$$

No they aren't independent and there does not exist any $f_{X,Y}(x,y)$ on $\Omega_{X,Y}$ such that X and Y are independent. This is because $\Omega_{X,Y}$ isn't a rectangle, which means it cannot be the Cartesian product of any ranges (and therefore the marginal ranges aren't independent).

6 Transformations of Random Variables

Suppose we had a random variable X and we knew its properties such as its cdf and pdf. Then, if we apply $Y = g(X)$, we would be interested in its distribution as well. We will show how to derive the distribution of Y if it is a function of X, and we will suppose they are both continuous.

Suppose $X \sim Unif(0,1)$. Find the distribution of $Y = -\ln X$. The general technique is to start with the cumulative distribution function.

We know that, for $x \in [0,1]$, $F_X(x) = x$ and so $P(X \geq x) = 1 - F_X(x) = 1 - x$.

$$F_Y(y) = P(Y \leq y) = P(-\ln X \leq y) = P(\ln X \geq -y) = P(X \geq e^{-y}) = 1 - e^{-y}$$

where the first equality is by definition of cdf, the second is by definition of Y, the third and fourth are algebra, and the last is by the fact about the uniform cdf, provided $e^{-y} \in [0,1]$, or equivalently, $y \in [0,\infty)$.

Now we can differentiate to get the density,

$$f_Y(y) = F_Y'(y) = e^{-y}, y \in [0,\infty)$$

It turns out that $Y \sim Exp(1)$! You can apply this technique in the more general case as well!

7 Conditional Expectation, Law of Total Expectation

Suppose X and Y are jointly distributed, and we'll suppose here that they are discrete random variables, but this applies in the more general case as well. We'll define the **conditional expectation of X given $Y = y$** as

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y)$$

We will do an example to illustrate. Suppose we have two fair four-sided dice, and let X be the value on the first, and Y the value on the second. Let $S = \max\{X, Y\}$, $T = \min\{X, Y\}$. What is the conditional expectation of $S|T = t$?

We start with the joint pmf from last time:

$p_{S,T}(s,t)$

$s\backslash t$	1	2	3	4
1	1/16	0	0	0
2	2/16	1/16	0	0
3	2/16	2/16	1/16	0
4	2/16	2/16	2/16	1/16

Let us find the conditional pmf of $S|T$. All we have to do is normalize so the columns sum to 1.

$p_{S|T}(s|t)$

$s\backslash t$	1	2	3	4
1	1/7	0	0	0
2	2/7	1/5	0	0
3	2/7	2/5	1/3	0
4	2/7	2/5	2/3	1

So now, to find

$$E[S|T = 4] = \sum_s s p_{S|T}(s|4) = 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot 1 = 4$$

This is completely expected! If the minimum value was 4, the maximum has no choice but to take on the value of 4.

$$E[S|T = 1] = \sum_s s p_{S|T}(s|1) = 1 \cdot \frac{1}{7} + 2 \cdot \frac{2}{7} + 3 \cdot \frac{2}{7} + 4 \cdot \frac{2}{7} = \frac{19}{7} \approx 2.714$$

One can easily calculate $E[S] = 3.125$. Notice that $E[S] = 3.125 > 2.714 = E[S|T = 1]$. Why does this make sense?

Now that we know what conditional expectation is, we have a very natural extension from the law of total probability called the **law of total expectation**. This says that

$$E[X] = \sum_y E[X|Y = y]P(Y = y) = \sum_y E[X|Y = y]p_Y(y)$$

Proof:

$$\sum_y E[X|Y = y]P(Y = y) = \sum_y \sum_x x p_{X|Y}(x|y)p_Y(y) = \sum_x x \sum_y p_{X,Y}(x, y)$$
$$= \sum_x x p_X(x) = E[X]$$

where the first equality is by definition of conditional expectation, the second is by definition of conditional probability, the third is by definition of marginal distribution, and the fourth is by definition of $E[X]$.

Q.E.D.