THE GAMMA AND CHI-SQUARE DISTRIBUTIONS*

Ewa Paszek

Abstract

This course is a short series of lectures on Introductory Statistics. Topics covered are listed in the Table of Contents. The notes were prepared by Ewa Paszek and Marek Kimmel. The development of this course has been supported by NSF 0203396 grant.

1 GAMMA AND CHI-SQUARE DISTRIBUTIONS

In the (approximate) Poisson process[1] with mean λ, we have seen that the waiting time until the first change has an exponential distribution[2]. Let now **W** denote the waiting time until the αth change occurs and let find the distribution of **W**. The distribution function of **W** ,when $w \geq 0$ is given by

$$F(w) = P(W \leq w) = 1 - P(W > w) = 1 - P(fewer\_than\_\alpha\_changes\_occur\_in\_[0,w])$$
$$= 1 - \sum_{k=0}^{\alpha-1} \frac{(\lambda w)^k e^{-\lambda w}}{k!},$$

since the number of changes in the interval $[0, w]$ has a Poisson distribution with mean λw. Because **W** is a continuous-type random variable, $F'(w)$ is equal to the p.d.f. of **W** whenever this derivative exists. We have, provided **w**>0, that

$$F'(w) = \lambda e^{-\lambda w} - e^{-\lambda w} \sum_{k=1}^{\alpha-1} \left[\frac{k(\lambda w)^{k-1}\lambda}{k!} - \frac{(\lambda w)^k \lambda}{k!} \right] = \lambda e^{-\lambda w} - e^{-\lambda w} \left[\lambda - \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!} \right]$$
$$= \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!} e^{-\lambda w}.$$

1.1 Gamma Distribution

Definition 1:
1. If $w < 0$, then $F(w) = 0$ and $F'(w) = 0$, a p.d.f. of this form is said to be one of the **gamma type**, and the random variable **W** is said to have **the gamma distribution**.

*Version 1.3: Oct 8, 2007 4:18 pm -0500
†http://creativecommons.org/licenses/by/2.0/
[1]"POISSON DISTRIBUTION", Definition 1: "POISSON PROCCESS" <http://cnx.org/content/m13125/latest/#def_1>
[2]"THE UNIFORM AND EXPONENTIAL DISTRIBUTIONS": Section An Exponential Distribution <http://cnx.org/content/m13128/latest/#sec_4>

2. The **gamma function** is defined by

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy, 0 < t.$$

This integral is positive for $0 < t$, because the integrand id positive. Values of it are often given in a table of integrals. If $t > 1$, integration of gamma fnction of **t** by parts yields

$$\Gamma(t) = \left[-y^{t-1} e^{-y} \right]_0^\infty + \int_0^\infty (t-1) y^{t-2} e^{-y} dy = (t-1) \int_0^\infty y^{t-2} e^{-y} dy = (t-1) \Gamma(t-1).$$

Example 1

Let $\Gamma(6) = 5\Gamma(5)$ and $\Gamma(3) = 2\Gamma(2) = (2)(1)\Gamma(1)$. Whenever $t = n$, a positive integer, we have, be repeated application of $\Gamma(t) = (t-1)\Gamma(t-1)$, that $\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)...(2)(1)\Gamma(1)$.

However,

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

Thus when **n** is a positive integer, we have that $\Gamma(n) = (n-1)!$; and, for this reason, the gamma is called **the generalized factorial**.

Incidentally, $\Gamma(1)$ corresponds to 0!, and we have noted that $\Gamma(1) = 1$, which is consistent with earlier discussions.

1.1.1 SUMMARIZING

The random variable **x** has **a gamma distribution** if its p.d.f. is defined by

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, 0 \le x < \infty. \quad (1)$$

Hence, **w**, the waiting time until the α th change in a Poisson process, has a gamma distribution with parameters α and $\theta = 1/\lambda$.

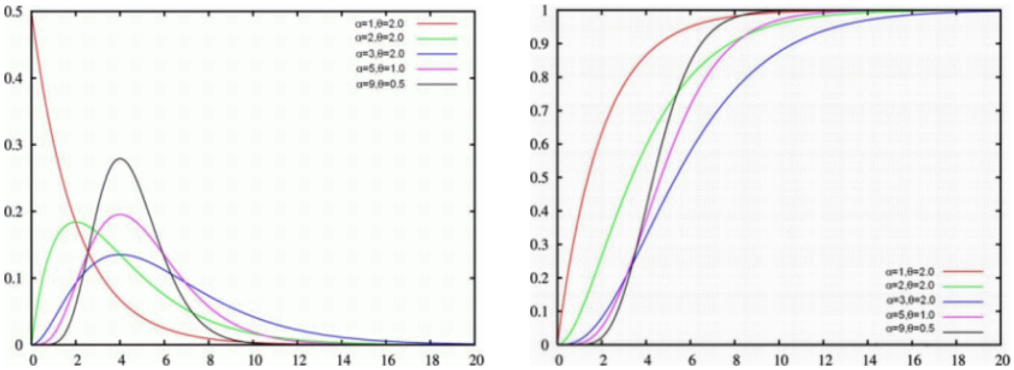
Function $f(x)$ actually has the properties of a p.d.f., because $f(x) \ge 0$ and

$$\int_{-\infty}^\infty f(x) dx = \int_0^\infty \frac{x^{\alpha-1} e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha} dx,$$

which, by the change of variables $y = x/\theta$ equals

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

The mean and variance are: $\mu = \alpha\theta$ and $\sigma^2 = \alpha\theta^2$.



(a) Gamma Distribution (b)

Figure 1: The p.d.f. and c.d.f. graphs of the Gamma Distribution. (a) The c.d.f. graph. (b) The p.d.f. graph.

1.1.2

Example 2

Suppose that an average of 30 customers per hour arrive at a shop in accordance with Poisson process. That is, if a minute is our unit, then $\lambda = 1/2$. What is the probability that the shopkeeper will wait more than 5 minutes before both of the first two customers arrive? If **X** denotes the waiting time in minutes until the second customer arrives, then **X** has a gamma distribution with $\alpha = 2, \theta = 1/\lambda = 2$. Hence,

$$p(X > 5) = \int_5^\infty \frac{x^{2-1}e^{-x/2}}{\Gamma(2)\,2^2}dx = \int_5^\infty \frac{xe^{-x/2}}{4}dx = \frac{1}{4}\left[(-2)\,xe^{-x/2} - 4e^{-x/2}\right]_5^\infty = \frac{7}{2}e^{-5/2} = 0.287.$$

We could also have used equation with $\lambda = 1/\theta$, because α is an integer

$$P(X > x) = \sum_{k=0}^{\alpha-1} \frac{(x/\theta)^k e^{-x/\theta}}{k!}.$$

Thus, with **x**=5, α=2, and $\theta = 2$, this is equal to

$$P(X > x) = \sum_{k=0}^{2-1} \frac{(5/2)^k e^{-5/2}}{k!} = e^{-5/2}\left(1 + \frac{5}{2}\right) = \left(\frac{7}{2}\right)e^{-5/2}.$$

1.2 Chi-Square Distribution

Let now consider the special case of the gamma distribution that plays an important role in statistics.

Definition 2:
Let **X** have a gamma distribution with $\theta = 2$ and $\alpha = r/2$, where **r** is a positive integer. If the p.d.f. of **X** is

$$f(x) = \frac{1}{\Gamma(r/2)\,2^{r/2}} x^{r/2-1} e^{-x/2}, 0 \le x < \infty. \quad (2)$$

We say that **X** has **chi-square distribution** with **r** degrees of freedom, which we abbreviate by saying is $\chi^2(r)$.

The **mean** and the **variance** of this chi-square distributions are

$$\mu = \alpha\theta = \left(\frac{r}{2}\right) 2 = r$$

and

$$\sigma^2 = \alpha\theta^2 = \left(\frac{r}{2}\right) 2^2 = 2r.$$

That is, the mean equals the number of degrees of freedom and the variance equals twice the number of degrees of freedom.

In the fugure 2 (Figure 2) the graphs of chi-square p.d.f. for **r**=2,3,5, and 8 are given.

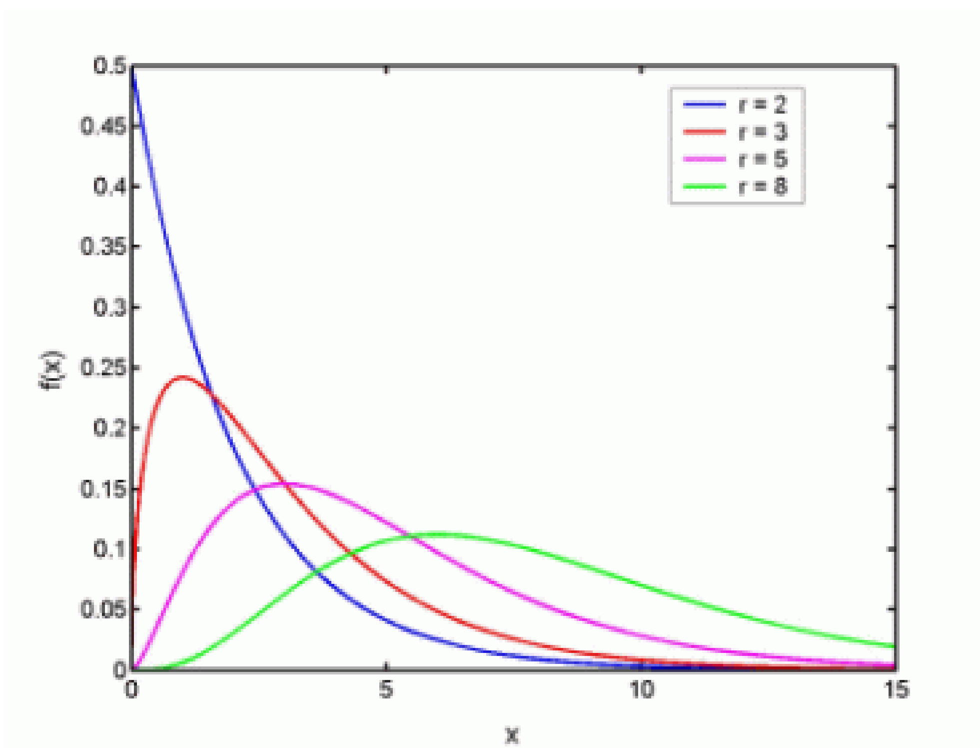


Figure 2: The p.d.f. of chi-square distribution for degrees of freedom **r**=2,3,5,8.

NOTE: the relationship between the mean $\mu = r$, and the point at which the p.d.f. obtains its maximum.

Because the chi-square distribution is so important in applications, tables have been prepared giving the values of the distribution function for selected value of **r** and **x**,

$$F(x) = \int_0^x \frac{1}{\Gamma(r/2)\, 2^{r/2}} w^{r/2-1} e^{-w/2} dw. \tag{3}$$

Example 3

Let **X** have a chi-square distribution with **r** =5 degrees of freedom. Then, using tabularized values,

$$P(1.145 \leq X \leq 12.83) = F(12.83) - F(1.145) = 0.975 - 0.050 = 0.925$$

and

$$P(X > 15.09) = 1 - F(15.09) = 1 - 0.99 = 0.01.$$

Example 4

If **X** is $\chi^2(7)$, two constants, **a** and **b**, such that $P(a < X < b) = 0.95$, are **a**=1.690 and **b**=16.01.

Other constants **a** and **b** can be found, this above are only restricted in choices by the limited table.

Probabilities like that in Example 4 (Example 4) are so important in statistical applications that one uses special symbols for **a** and **b**. Let α be a positive probability (that is usually less than 0.5) and let **X** have a chi-square distribution with **r** degrees of freedom. Then $\chi^2_\alpha(r)$ is a number such that $P\left[X \geq \chi^2_\alpha(r)\right] = \alpha$

That is, $\chi^2_\alpha(r)$ is the 100(1-α) percentile (or upper 100a percent point) of the chi-square distribution with **r** degrees of freedom. Then the 100α percentile is the number $\chi^2_{1-\alpha}(r)$ such that $P\left[X \leq \chi^2_{1-\alpha}(r)\right] = \alpha$. This is, the probability to the right of $\chi^2_{1-\alpha}(r)$ is 1-α. SEE fugure 3 (Figure 3).

Example 5

Let **X** have a chi-square distribution with seven degrees of freedom. Then, using tabularized values, $\chi^2_{0.05}(7) = 14.07$ and $\chi^2_{0.95}(7) = 2.167$. These are the points that are indicated on Figure 3.

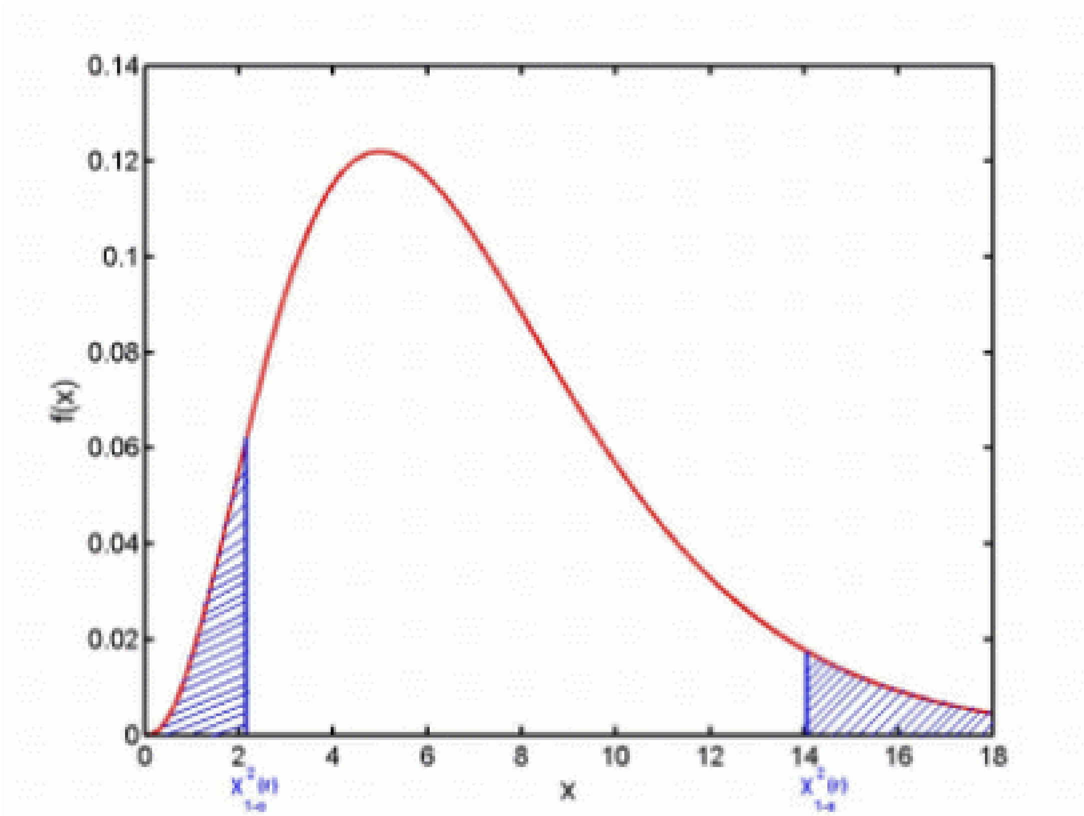


Figure 3: $\chi^2_{0.05}(7) = 14.07$ and $\chi^2_{0.95}(7) = 2.167$.

2