Course:	Mathematical Statistics
Term:	Fall 2017
Instructor:	Gordan Žitković

Lecture 7

The normal, $\chi^2(n)$ and the Gamma distributions

7.1 The normal distribution

Definition, the pdf and the mgf of the normal distribution. First, we collect the useful facts about the normal distribution. Most of them have already been mentioned in the previous lectures, but some will be new. As you remember, the **normal distribution (family)** is a continuous probability distribution, parametrized by two parameters, $\mu \in \mathbb{R}$ and $\sigma > 0$, with the pdf

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

The cdf does not have a closed form, but the mgf does:

$$m_Y(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

The Central Limit Theorem and the approximation of the binomial distribution. One of the reasons for the ubiquity of the normal distribution is that it serves as an approximation to the binomial distribution. More generally, it serves as an approximation to a (properly normalized) sum of independent and identically distributed random variables - this result is known as the Central Limit Theorem:

Theorem 7.1.1 (Central Limit Theorem)**.** *Let $Y_1, Y_2, \dots, Y_n, \dots$ be a sequence of independent random variables with the same distribution. If $\mu = \mathbb{E}[Y_i]$ and $\sigma^2 = \operatorname{Var}[Y_i] < \infty$, then the distribution of the normalized sum*

$$\frac{S_n - \mathbb{E}[S_n]}{\operatorname{sd}[S_n]} \text{ where } S_n = Y_1 + \dots + Y_n$$

converges[1] towards the unit normal $N(0,1)$.

[1] this kind of convergence is called the **weak convergence** or **convergence in distribution** and is very important in probability theory. We will not talk about it in this class; the reader should simply think of $(S_n - \mathbb{E}[S_n]) / \operatorname{sd}[S_n]$ for n large as being close to the normal $N(0,1)$ distribution as far as any practical computation involving probabilities or expectations is concerned.

The Central Limit Theorem is used in many ways. One of the most important for the foundations of statistics is the following. Let Y be a binomially distributed random variable with parameters n and p, i.e., $Y \sim b(n,p)$. We can represent the distribution of Y as a sum of n independent $B(p)$ random variables - let us call them $Y_1, Y_2, \dots$. Let us apply the Central Limit Theorem to that sequence. Since $Y = S_n$, it allows us to conclude that the random variable

$$\frac{S_n - \mathbb{E}[S_n]}{\mathrm{sd}[S_n]} \text{ which, in this case, equals } \frac{Y-np}{\sqrt{np(1-p)}},$$

is "close" to the normal $N(0,1)$ distribution. This is the well-known **normal approximation to the binomial distribution.** The word "close" does not mean anything, really, and one needs a more precise measure of the quality of this approximation. There are many mathematical theorems written about it, and the ultimate answer will depend on the level of precision your calculation requires, but the following rule seems to work well in most applications:

> **A rule of thumb:** *the normal approximation to the binomial $b(n,p)$ can be used if*
>
> $$np > 10 \text{ and } n(1-p) > 10.$$

What can one do when $np < 10$ or $n(1-p) < 10$? In some cases, one should use the Poisson approximation (coming from a different version of the Central Limit Theorem). Here is another practical prescription (due to Jay Devore):

> **Another rule of thumb:** *the binomial $b(n,p)$ is well approximated by the Poisson distribution with parameter $\lambda = np$ if*
>
> $$n > 50 \text{ and } np < 5.$$

This class will focus only on the cases where the normal approximation is appropriate. That is not to say that the Poisson approximation is not important in statistics, too, but we will not use it in these notes.

Example 7.1.2. A basketball player hits a free throw with probability $p = 3/4$, and the outcomes of different throws can be considered independent of each other. What is the probability that she will hit at least 160 out of 200 free throws?

The number of hits Y is binomially distributed with parameters $n = 432$ and $p = 3/4$. The probability we are interested in is $\mathbb{P}[Y \geq 333]$ and, using the pmf of a binomial distribution, it can be represented as

$$\mathbb{P}[Y \geq 333] = \sum_{k=333}^{432} \mathbb{P}[Y = k] = \sum_{k=333}^{432} \binom{432}{k} \left(\tfrac{3}{4}\right)^k \left(\tfrac{1}{4}\right)^{432-k}. \quad (7.1.1)$$

This is as far as it gets, though. This sum is difficult to evaluate - it has 41 terms, and each of them involves a binomial coefficient and large powers of small numbers, such as $(1/4)^{432}$. Luckily, $np = 324$ and $n(1-p) = 108$, so our rule of thumb allows us to use the normal approximation:

$$\text{the distribution of } \frac{Y - \mathbb{E}[Y]}{\text{sd}[Y]} \text{ is approximately } N(0,1).$$

Since $\mathbb{E}[Y] = 432 \times 3/4 = 324$ and $\text{sd}[Y] = 9$, the probability $\mathbb{P}[Y \geq 333]$ is the same as the probability

$$\mathbb{P}[\tfrac{Y-324}{9} \geq \tfrac{333-324}{9}] = \mathbb{P}[\tfrac{Y-324}{9} \geq 1].$$

Since $\frac{Y-324}{9}$ is well approximation by a normal $N(0,1)$-distribution, this probability is well approximated by the probability

$$\mathbb{P}[Z \geq 1],$$

where $Z \sim N(0,1)$. At this point, one needs to use software (or distribution tables) to find out that $\mathbb{P}[Z \geq 1] \approx 0.1586$. If we compute the sum from (7.1.1) exactly, we get 0.1727. This is not bad, but it is not great either. It can be improved significantly using the so-called *continuity correction*[1], but we will not need it in this class.

[1] With the continuity correction, we would get 0.1724 instead of 0.1586 in this case. That would decrease the relative error from about 8% to 0.1% !

Just to get a flavor of how a Poisson approximation might work, here is another example

Example 7.1.3. A new addition of Kafka's "Metamorphosis" has 72 pages. The printing press often malfunctions and introduces typos. The probability that a word will contain a typo is $p = 10^{-4}$, and we can assume there are 250 words per page. A printed book gets thrown away if it contains more than 5 typos. What is the probability that a book will get thrown away?

We are talking about a random variable Y with the binomial distribution with parameters $n = 72 \times 250 = 18,000$ words and $p = 10^{-4}$. The book will get thrown away if $Y > 5$. Since $np = 1.8$, the normal approximation might not be the best idea. On the other hand, $n > 50$ and $np < 5$, so we can use the Poisson approximation. More precisely, the probability $\mathbb{P}[Y > 20]$ is well approximated by the probability

$\mathbb{P}[W > 20]$, where $W \sim P(\lambda)$ for $\lambda = np = 1.8$. Then,

$$\begin{aligned}\mathbb{P}[W > 5] = 1 - (&\mathbb{P}[W = 0] + \mathbb{P}[W = 1] + \mathbb{P}[W = 2] + \mathbb{P}[W = 3]+ \\ &+ \mathbb{P}[W = 4] + \mathbb{P}[W = 5]) \\ &= 1 - e^{-1.8}\Big(1 + \tfrac{1.8}{1} + \tfrac{(1.8)^2}{2!} + \tfrac{(1.8)^3}{3!} + \tfrac{(1.8)^4}{4!} + \tfrac{(1.8)^5}{5!}\Big) \\ &\approx 1 - 0.1653 \times 5.9869 = 0.01037.\end{aligned}$$

Therefore, about 1% of the books will get thrown away.

Sums of independent normals. Let $Z_1, Z_2, \dots, Z_n$ be independent and normally distributed, but with possibly different parameters, $\mu_1, \dots, \mu_n$, and $\sigma_1, \dots, \sigma_n$. The sum

$$Z = Z_1 + Z_2 + \dots + Z_n$$

appears in statistics very often. In the section on the moment-generating functions we stated that the mgf of the sum of independent random variables is the product of individual mgfs. The conclusion of Example **??** of that section - which used the fact that $m_{Z_i}(t) = \exp(\mu_i t + \frac{1}{2}\sigma_i^2 t^2)$ - was that

$$m_Z(t) = \exp\Big((\mu_1 + \dots + \mu_n)t + \tfrac{1}{2}(\sigma_1^2 + \dots + \sigma_n^2)t^2\Big),$$

which can be readily recognized as the mgf of the normal distribution with mean $\mu = \mu_1 + \dots + \mu_n$ and $\sigma = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$. This fact is so important that we restate it as a proposition:

Proposition 7.1.4. *Let $Z_1, Z_2, \dots, Z_n$ be independent random variables, with $Z_i \sim N(\mu_i, \sigma_i)$. Then their sum $Z = Z_1 + \dots + Z_n$ is also normally distributed, with parameters $\mu = \mu_1 + \dots + \mu_n$ and $\sigma = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$.*

Proposition 7.1.4 above did not need to tell us what μ and σ are. Simply knowing that Z is normally distributed, we can recover the parameters μ and σ as its mean $\mathbb{E}[Z]$ and its standard deviation $\mathrm{sd}[Z]$ so that

$$\mu = \mathbb{E}[Z] = \mathbb{E}[Z_1 + \dots + Z_n] = \mathbb{E}[Z_1] + \dots + \mathbb{E}[Z_n] = \mu_1 + \mu_2 + \dots + \mu_n,$$

and

$$\sigma^2 = \mathrm{Var}[Z] = \mathrm{Var}[Z_1] + \dots + \mathrm{Var}[Z_n] = \sigma_1^2 + \dots + \sigma_n^2.$$

The beautiful conclusion of the Central Limit Theorem is that the sum $Y = Y_1 + \dots + Y_n$ is still approximately normal, even if Y_1, Y_2, $\dots Y_n$ are *not normally distributed* themselves. They can have (practically) any distribution they want, as long as they are independent, and their means $\mu_i = \mathbb{E}[Y_i]$ and standard deviations $\sigma_i = \mathrm{sd}[Y_i]$ are finite for each i. This is one of the main reasons why statistics works so well!

7.2 The $\chi^2(n)$-distribution

When we talked about functions of random variables in Lecture 4 before, one of the examples involved computing the pdf of the square $W = Y^2$, where $Y \sim N(0,1)$. Even though the h-method could not have been used, the cdf-method made it possible to derive the following pdf for W:

$$f_W(w) = \tfrac{1}{\sqrt{2\pi w}} e^{-w/2} \mathbf{1}_{\{w>0\}}. \tag{7.2.1}$$

We also remarked that the distribution with that particular pdf is called the χ^2**-distribution**. Problem 7.6.8 of Lecture 6 lead us through the computation of its mgf:

$$m_W(t) = (1-2t)^{-1/2}.$$

It went on to identify the distribution of the sum $Z_1^2 + Z_2^2$ of squares of two independent unit normals. We can continue adding more terms:

Definition 7.2.1. The χ^2**-distribution** with n **degrees of freedom** is the distribution of a sum

$$W = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

of squares of n independent, unit normal ($N(0,1)$) random variables Z_i. We denote this by $W \sim \chi^2(n)$.

We know already that $\chi^2(1)$ is the χ^2-distribution with pdf given by (7.2.1). It is shown in Problem 7.6.8 that $\chi^2(2)$ is nothing but the exponential $E(2)$ distribution. Unfortunately, no such nice identification can be made for larger n (at least not in the family of the named distributions we talked about in this class). Since the mgf of $\chi^2(1)$ is $(1-2t)^{-1/2}$, it follows immediately from the definition that the mgf of a $\chi^2(n)$ distribution is

$$m_W(t) = (1-2t)^{-1/2} \times (1-2t)^{-1/2} \times \dots \times (1-2t)^{-1/2} = (1-2t)^{-n/2}.$$

It is interesting to note that we could derive the mgf of W without ever mentioning its pdf, cdf, or any other characteristic. In fact, we do not at this point even know whether it is discrete, continuous or neither. In turns out (but we do not prove it here) that W is a continuous random variable and that its pdf is given by a (relatively) simple formula, but we will not need it in these notes (see Problem 7.6.11 below if you are dying to know what the pdf is).

The knowledge of the mgf gives us plenty of information about W. If we differentiate its mgf several times and plug in $t = 0$ we obtain

$$\mu'_k = \mathbb{E}[W^k] = n(n+2)(n+4)\dots(n+2k).$$

In particular,

$$\mathbb{E}[W] = n \text{ and } \operatorname{Var}[W] = 2n.$$

The terminology *degrees of freedom* comes from statistics (linear regression). It refers, loosely, to the number of independent normals used to build W, and we will have more to say about it in a few lectures.

7.3 The gamma distribution

The χ^2-distribution comes with a single parameter n - the number of degrees of freedom, but does not allow for scaling: if $Y \sim \chi^2(n)$, then $5Y$ is not $\chi^2(m)$-distributed for any m. In order to remedy this situation we can add another parameter to the χ^2-family, and, while we are at it, we cal also allow for the number of degrees of freedom n to take any positive (possibly non-integer) value. The easiest way to accomplish this is to use the fact that $m_{aY}(t) = m_Y(at)$ and use it with the mgf of the χ^2-distribution:

Definition 7.3.1. A random variable Y is said to have the **gamma distribution** with parameters $k > 0$ and $\tau > 0$ - denoted by $\Gamma(k, \tau)$ - if its moment-generating function is given by

$$m_Y(t) = (1 - \tau t)^{-k}.$$

The parameter k is usually called the **shape parameter** and τ is the **scale parameter**. Like the χ^2-distributions, the gamma distribution takes only non-negative values. By a simple differentiation of its mgf, we obtain the following values for the mean and the variance of the $\Gamma(k, \tau)$-distribution:

$$\mathbb{E}[Y] = k\tau \text{ and } \operatorname{Var}[Y] = k\tau^2.$$

As we already mentioned, the gamma family contains all exponential and $\chi^2(n)$-distributions:

Exponential distribution	$E(\tau)$	$\Gamma(1, \tau)$
χ^2 distribution	$\chi^2(n)$	$\Gamma(\frac{n}{2}, 2)$

It is also clear from the form of the mgf of the gamma distribution is that if $Y_1 \sim \Gamma(k_1, \tau)$ and $Y_2 \sim \Gamma(k_2, \tau)$ (note, τ must be the same!) and Y_1 and Y_2 are independent then

$$Y_1 + Y_2 \sim \Gamma(k_1 + k_2, \tau). \tag{7.3.1}$$

Example 7.3.2. The time between two successive cars passing a given point on a highway can be modeled by an exponential distribution with mean $\tau = 2$ (seconds). Assuming that the times between cars are independent, what is the distribution of the time it takes for exactly 100 cars to pass?

Let Y_1 be the time between the start of the experiment and the moment car 1 passes the point, Y_2 between the cars 1 and 2, etc. We are looking for the distribution of $Y = Y_1 + \cdots + Y_{100}$. Each Y_1 is an $E(2)$ random variable, which is also a gamma distribution with parameters $k = 1$ and $\tau = 2$. Thanks to (7.3.1), the sum, Y, has the $\Gamma(100, 2)$ distribution.

Like in the case of the $\chi^2(n)$-distribution, we never mentioned the pdf (or the cdf, or ...) of a Gamma distribution. The reason is that we do not really need it, and it involves the so-called Gamma function which is not familiar to most students. Those who are curious should try problem 7.6.11. Also, here are some plots of the gamma pdf for representative values of the shape parameter k (we keep $\tau = 1$; to vary τ, simply stretch or compress the horizontal axes by the factor of τ, and do the opposite to the vertical axes):

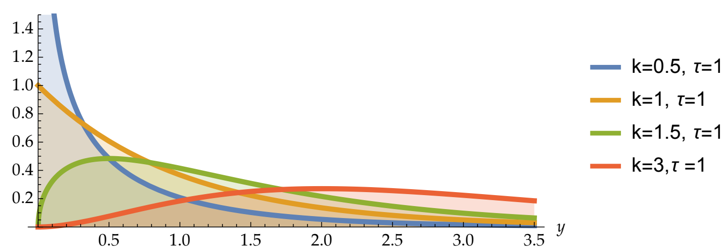

Figure 1. Pdf of the gamma distribution with $\tau = 1$ and $k = 0.5, 1, 1.5$ and $k = 3$.

7.4 Computations

We introduced the gamma distribution without specifying what its pdf is, and mentioned that it did not really matter. But how do we compute the probabilities of expectations then? If we had the pdf, we would simply integrate it to get the result, but no such a procedure is available using only the mgf. The truth is that, even with the access to pdfs these integrals cannot be evaluated explicitly, and we need to resort to various numerical procedures. We do not discuss these procedures at all, other than saying that they are built on interesting mathematics, and are mostly very quick and accurate. We do give a short list of R (S-Plus) commands corresponding to some of the most important distributions and some of the most important probabilities. These commands as well as the arguments are composed of two part. The first one - called the **root** - is a single letter and it stands for the kind of output one is interested in

initial letter	output	first argument
`p`	cdf	`(y,`
`d`	pdf or pmf	`(y,`
`q`	quantile	`(q,`

The second part of the command - called the **tag** - describes the distribution and the rest of the arguments describe the parameters of the distribution:

tag	distribution	other arguments	note
`binom`	binomial	`n,p)`	
`chisq`	χ^2	`n)`	
`exp`	exponential	`r)`	$r = 1/\tau$
`gamma`	gamma	`k,r)`	$r = 1/\tau$
`norm`	normal	μ,σ`)`	
`pois`	poisson	λ`)`	
`unif`	uniform	`l,r)`	

Example 7.4.1. Suppose that we want to know the probability $\mathbb{P}[Y \leq 20]$, where Y is a normal random variable with $\mu = 0.2$ and $\sigma = 13$. We would then construct the appropriate command by combining the root `p` (because we need the cdf), with the tag `norm`; the arguments would be $(20, 0.2, 13)$. In a real-life R-session it looks like this

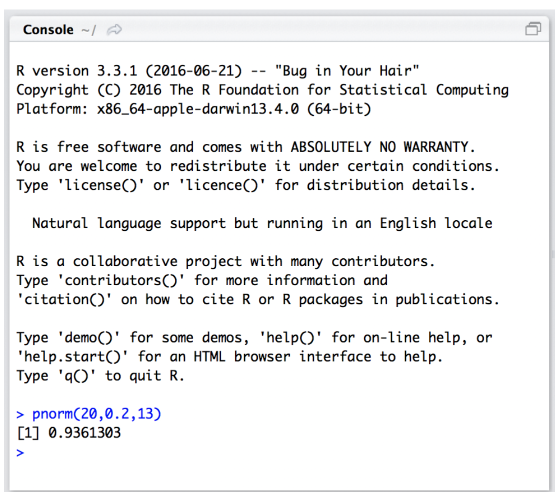

Figure 2. An R session in which we compute the probability $\mathbb{P}[Y \leq 20]$, where $Y \sim N(0.2, 13)$.

Therefore, $\mathbb{P}[Y \leq 20] \approx 0.936$.

If we wanted the 75% quantile of the exponential distribution with $\tau = 2$, we would use the root `q`, the rest `exp`, the first argument 0.75 and the second argument 0.5 (because R parameterizes the exponential and gamma distributions differently). Again, in real life:

```
Console ~/
> qexp(0.75,0.5)
[1] 2.772589
>
```

Figure 3. An R session in which we compute the 0.75-quantile of the exponential distribution with parameter (mean) $\tau = 2$

which makes the value of the desired quantile about 2.773. Let us try and check if this is indeed the 0.75-th quantile, i.e., if $F_Y(2.773) = 0.75$, where F_Y is the cdf of the exponential $E(2)$. We can do it using R:

```
Console ~/
> pexp(2.773,0.5)
[1] 0.7500514
>
```

Figure 4. An R session in which we check that 2.773 is indeed the (approximate) 0.75-quantile of the exponential distribution with parameter (mean) $\tau = 2$

We will not use R in this class, so the way we use the R commands described above is as shorthands and alternatives to, by now really outdated, tables. Here is an example:

Example 7.4.2. Let Z_1, Z_2 and Z_3 be three independent unit normal ($N(0,1)$) random variables. If (Z_1, Z_2, Z_3) are interpreted as the coordinates of a random point in the space, what is the probability that the distance from (Z_1, Z_2, Z_3) to the origin is at least 2 ?

The distance R to the origin is given by

$$R = \sqrt{Z_1^2 + Z_2^2 + Z_3^2},$$

and, so, the require probability is $\mathbb{P}[R \geq 2] = \mathbb{P}[R^2 \geq 4]$. Being a sum of squares of three independent normals, R^2 has the $\chi^2(3)$ distribution, and, so,

$$\mathbb{P}[R^2 \geq 4] = 1 - \mathbb{P}[R^2 < 4] = 1 - \texttt{pchisq(4,3)}.$$

We don't need to start an R-session here; leaving the answer like this is perfectly fine as we know exactly what to do if we needed a numerical answer.

7.5 Simulation

In addition to computations of probabilities and expected values of various quantities related to random variables with different distributions, (statistical) software allows us to **simulate** these quantities, as well. Simulation from a given distribution, roughly speaking, refers to an algorithmic production of a sequence of numbers which "looks like" a sequence of independent draws from that distribution would have. Instead of actually tossing an unbiased coin 10000 times, it is much more efficient to write code that produces a sequence of 10000 symbols, each from the set $\{H, T\}$, that resembles an actual sequence of coin tosses. How something like that is actually achieved, and whether it is even possible to produce truly random numbers by algorithmic means, is a deep subject which we do not even touch here. Instead, we focus on the mechanics of it and the appropriate R commands.

The only thing to do, really, is to add a new root namely `r`, i.e., to add another row to the table of possible roots (and keep all the already described tags):

initial letter	output	first argument
`r`	random numbers	`(n,`

The argument n tells us how many "draws" from the distribution specified in the tag portion we need.

Example 7.5.1. If we want to take 5 draws from the uniform distribution on the interval $(2, 3)$, we would issue the command `runif(5,2,3)`:

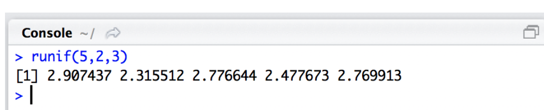

Figure 5. An R session in which we take 5 draws from the uniform distribution $U(2, 3)$

If we issued the same command again, we would get a different sample:

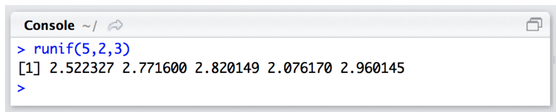

Figure 6. An R session in which we take another 5 draws from the uniform distribution $U(2,3)$

Here is what happens when I draw many (100) samples from the uniform distribution:

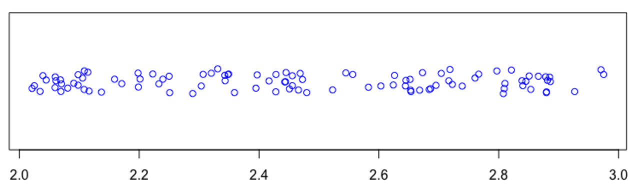

Figure 7. 100 draws from the uniform distribution U(2,3). The y-coordinates of points are irrelevant - they are randomly "jittered" to avoid overlap.

Here is another run of the same R command

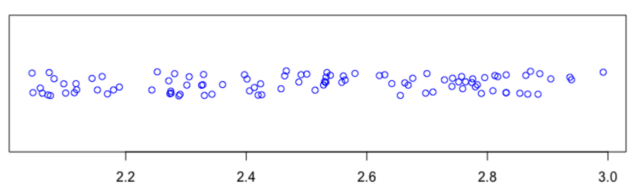

Figure 8. 100 more draws from the uniform distribution U(2,3).

Let us try a different distribution, say exponential with parameter 2. The R command for 5 draws from $E(2)$ is `rexp(5,0.5)` (remember, R uses $1/\tau$ and not τ as the parameter):

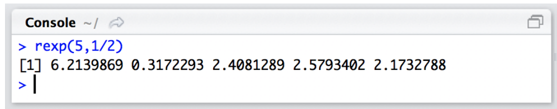

Figure 9. An R session in which we take 5 draws from the exponential distribution with parameter $\tau = 2$.

100 draws from $E(2)$, will look like this:

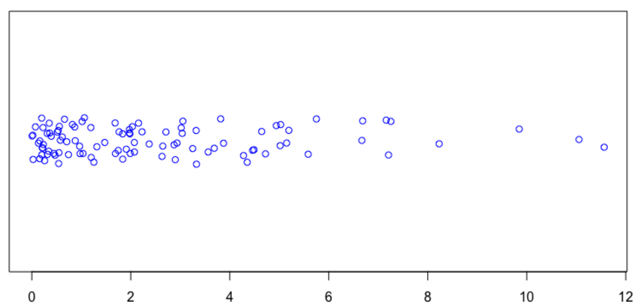

Figure 10. 100 draws from the exponential distribution $E(2)$.

7.6 Problems

Problem 7.6.1. A 13-sided die is thrown 169 times (each of the numbers $1, 2, \dots, 13$ is equally likely on each throw). Every time 5 or a larger number is obtained, the player wins a candy bar. Every time she gets a 13, she gets to pick a card from a deck of 52. If the picked card is an ace (there are 4 aces in the deck), she gets a free-massage coupon. What is the probability that the player will receive

1. at least 105 candy bars?
2. at least one free-massage coupon. (*Hint:* Check the rules of thumb for the normal and the Poisson approximations.)

(*Note:* In the first part, write your answer using an appropriate R command. If you have easy access to R, run it to see what the numerical value is. If you don't, just leave it.)

Problem 7.6.2. The best approximation (among the offered answers) to the binomial distribution with $n = 100$ and $p = 0.2$ is

(a) $N(0,1)$ (b) $P(20)$ (c) $B(0.2)$ (d) $N(20,4)$ (e) $E(20)$

Problem 7.6.3. Let $Y_1, \dots, Y_{100}$ be independent random variables with the Bernoulli $B(p)$ distribution, with $p = 0.2$ The best approximation to $\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$ (among the offered answers) is

(a) $N(0,1)$ (b) $N(100,20)$ (c) $N(0.2,0.04)$ (d) $N(20,4)$ (e) $N(20,20)$

(*Note:* In our notation $N(\mu,\sigma)$ means normal with mean μ and *standard deviation* σ.)

Problem 7.6.4. Let $Y_1, \dots, Y_n$ be independent random variables with the $N(\mu,\sigma)$-distribution. Which of the following is the distribution of

$$\left(\tfrac{Y_1-\mu}{\sigma}\right)^2 + \left(\tfrac{Y_2-\mu}{\sigma}\right)^2 + \cdots + \left(\tfrac{Y_n-\mu}{\sigma}\right)^2 \;?$$

(a) $N(\mu,\sigma)$ (b) $N(0,1)$ (c) $\chi^2(1)$ (d) $\chi^2(n)$ (e) none of the above

Problem 7.6.5. Let Y_1, Y_2 be independent random variables with distribution $N(0,\sigma)$ where σ is the standard deviation. The distribution of $Y_1^2 + Y_2^2$ is

(a) $N(0,\sigma)$ (b) $\chi^2(2)$ (c) $\chi^2(2\sigma^2)$ (d) $E(2\sigma^2)$ (e) none of the above

Problem 7.6.6. You buy a box of 25 light bulbs. Once the first one burns you replace it with the next one, etc. If each bulb lasts an exponentially distributed amount of time with mean $\tau = 1$ (months), and there is no dependence between the bulbs, then the distribution of the time it will take to go through the whole box is

(a) $E(25)$ (b) $E(1/25)$ (c) $\Gamma(1,25)$, i.e., $k = 1$, $\tau = 25$
(d) $\Gamma(25,1)$, i.e., $k = 25$, $\tau = 1$ (e) $\chi^2(25)$

Problem 7.6.7. Let Y_1 and Y_2 be independent standard normal ($N(0,1)$) random variables. Then, in R notation, $\mathbb{P}[1 \le Y_1^2 + Y_2^2 \le 2]$ equals

(a) `pchisq(2,2)-pchisq(1,2)`

(b) `pnorm(2,2,1)`^{2} - `pnorm(1,1,1)`^{2}

(c) `dnorm(2,2,1)`^{2} - `dnorm(1,1,1)`^{2}

(d) `pnorm(`$\sqrt{2}$`,2,1)` - `pnorm(1,2,1)`

(e) none of the above

Problem 7.6.8 (The mgf of the χ^2 distribution)**.** We learned in class that the distribution of the random variable $Y = Z^2$, where Z is a standard normal, is called the χ^2**-distribution**. We also computed its pdf using the cdf method. The goal of this exercise is to compute its mgf.

1. Compute $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\beta x^2}\, dx$, where $\beta > 0$ is a constant. (*Hint:* Use the fact that $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(y-\mu)^2/\sigma^2}\, dy = 1$, as the expression inside the integral is the pdf of $N(\mu, \sigma)$. Plug in $\mu = 0$ and an appropriate value of σ.)
2. Compute the mgf of a χ^2-distributed random variable by computing $\mathbb{E}[e^{tZ^2}]$, where Z is a standard normal, i.e., $Z \sim N(0,1)$. (*Hint:* Compute this expectation by integrating the function $g(z) = e^{tz^2}$ against the standard normal density, and not the function $g(y) = e^{ty}$ against the density of Y, which we derived in class. It is easier this way.)
3. Let Z_1 and Z_2 be two independent standard normal random variables. Compute the mgf of $W = Z_1^2 + Z_2^2$, and then look through the notes for a (named) distribution which has the same mgf. What is it? What are its parameter(s)? (*Note:* If Z_1 and Z_2 are independent, then so are Z_1^2 and Z_2^2.)

Problem 7.6.9. Let $Y_1, \dots, Y_n$ be independent exponentially ($E(\tau)$)-distributed random variables with parameter $\tau > 0$. The distribution of $\frac{1}{\tau}(Y_1 + \dots + Y_n)$ is

(a) $\Gamma(n, 1)$ (b) $E(n/\tau)$ (c) $E(n\tau)$ (d) $\chi^2(n)$ (e) none of the above

where, $\Gamma(n, 1)$ means $\Gamma(k, \tau)$ with $k = n$ and $\tau = 1$.

Problem 7.6.10. A bottle is filled with gas. We think of it as a collection of n particles, each moving independently of others, with velocities modeled by random variables. The three components $V_1^{(n)}$, $V_2^{(n)}$ and $V_3^{(n)}$ of each particle's velocity $V^{(n)} = (V_1^{(n)}, V_2^{(n)}, V_3^{(n)})$ are also independent of each other and normally distributed with mean 0 and standard deviation σ.

1. The **kinetic energy** of a particle is given by

$$E_k = \tfrac{1}{2} m \Big((V_1^{(n)})^2 + (V_2^{(n)})^2 + (V_3^{(n)})^2 \Big),$$

where m is the particle's mass (we assume that all particles are identical). What is the distribution of the kinetic energy E_k of each individual particle?

2. The **total energy** of the gas is sum of kinetic energies of individual particles, i.e.,

$$E_t = \tfrac{1}{2}m \sum_{k=1}^{n} \Big((V_1^{(k)})^2 + (V_2^{(k)})^2 + (V_3^{(k)})^2 \Big).$$

   What is the distribution of the total energy E_t of the entire gas (n particles).

Problem 7.6.11. (*) Going from the pdf to the mgf is a matter of evaluating an integral. The opposite way is typically much harder. Thanks to the uniqueness theorem, one possibility is to guess and verify. Based on the pdf known in the special cases

$$\tfrac{1}{\tau} e^{-y/\tau} \mathbf{1}_{\{y>0\}} \text{ for } E(\tau) = \Gamma(1, \tau),$$

and

$$\tfrac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} \mathbf{1}_{\{y>0\}}, \text{ for } \chi^2 = \Gamma(\tfrac{1}{2}, 2),$$

an educated guess about the pdf of the $\Gamma(k, \tau)$-distribution would be

$$f_Y(y) = c y^{k-1} e^{-y/\tau} \mathbf{1}_{\{y>0\}}, \tag{7.6.1}$$

for some constant c which will guarantee that $\int_{-\infty}^{\infty} f_Y(y)\, dy = 1$. While this constant cannot be computed analytically for all k and τ, it helps if we introduce the **Gamma function**

$$\Gamma(k) = \int_0^\infty y^{k-1} e^{-y}\, dy, \; k > 0.$$

1. Use integration by parts to compute the values $\Gamma(1)$, $\Gamma(2)$, $\Gamma(3)$ and $\Gamma(4)$. Can you spot a pattern? Can you write $\Gamma(n+1)$ in terms of $\Gamma(n)$? What is $\Gamma(n)$ for $n \in \mathbb{N}$?
2. Express the constant c form (7.6.1) in terms of τ and $\Gamma(k)$.
3. Compute the mgf of the random variable whose pdf is given by (7.6.1), with the value of c you determined in 2. above. Was our guess correct?

Problem 7.6.12. (*) Let $Y_1, Y_2, \dots, Y_n$ be independent random variables each with the Bernoulli $B(p)$ distribution, for some $p \in (0,1)$.

1. Show that the mgf $m_{W_n}(t)$ of the random variable

$$W_n = \tfrac{Y_1 + Y_2 + \dots + Y_n - np}{\sqrt{np(1-p)}},$$

   can be written in the form

$$m_{W_n}(t) = (p e^{\frac{t}{\sqrt{n}}\alpha} + (1-p) e^{-\frac{t}{\sqrt{n}}\alpha^{-1}})^n, \tag{7.6.2}$$

   for some α and find its value.

2. Write down the Taylor approximations in t around 0 for the functions $\exp(\frac{t}{\sqrt{n}}\alpha)$ and $\exp(-\frac{t}{\sqrt{n}}\alpha^{-1})$, up to and including the term involving t^2. Then, substitute those approximations in (7.6.2) above. What to you get? When n is large, $\frac{t}{\sqrt{n}}\alpha$ and $\frac{t}{\sqrt{n}}\alpha^{-1}$ are close to 0 and it can be shown that the expression you got is the limit of $m_{W_n}(t)$, as $n \to \infty$.

3. What distribution is that limit the mgf of?

(*Note:* Convergence of mgfs corresponds to a very important mode of convergence, called the *weak convergence*. We will not talk about it in this class, but it is exactly the kind of convergence that appears in the central limit theorem, which is, in turn, behind the effectiveness of the normal approximation to binomial random variables. In fact, what you just did is a fundamental part of the proof of the central limit theorem.)