Department of Statistics, Yale University
STAT242b Theory of Statistics

Suggested Solutions to Homework 1

Compiled by Marco Pistagnesi

Problem 2.9

For the exponential, we have that the pdf is $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \{x > 0\}$, hence we have:

$F(x) = \int\limits_0^x \frac{1}{\beta} e^{-\frac{t}{\beta}} dt$, which is a trivial integral and can be evaluated in a single step. A rigorous way would be:

change the variable as: $y = \frac{t}{\beta} \Rightarrow \frac{dy}{dt} = \frac{1}{\beta} \Rightarrow dt = \beta dy$. Hence, taking care of the change in the limits too, we get:

$\frac{1}{\beta} \int_0^{\frac{x}{\beta}} e^{-y} \beta dy = \int_0^{\frac{x}{\beta}} e^{-y} dy = \left[-e^{-y}\right]_0^{\frac{x}{\beta}} = (1 - e^{-\frac{x}{\beta}}) 1 \{x > 0\}$,

where $1\{.\}$ is the indicator function, that takes the value 0 for $x < 0$. This is an important subtlety as the cdf is defined over the entire real line, no matter what is the range of x. We than have to specify that it will take the value 0 for $x < 0$ as the probability that $x < 0$ is null in this case.

Finding the inverse is also a trivial task, we can set the equation $1 - e^{-\frac{x}{\beta}} = y$ and solve for x. The result is $F^{-1}(y) = -\beta \log(1 - y)$.

Problem 3.19

The density of the the $U(0,1)$ is simply $f_X(x) = 1\{0 \leq x \leq 1\}$ and the graph is pretty easy... Then we know that for iid random variables :

$E(\overline{x}) = \mu$ and $Var(\overline{x}) = \frac{\sigma^2}{n}$, where μ and σ^2 are the mean and variance of any of the x_i's.

For the $U(0,1)$ we know that $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$ so we get $E(\overline{x}) = \frac{1}{2}$ and $Var(\overline{x}) = \frac{1}{12n}$. The plots, that I omit, indicate us that the expectation is constant while the variance decreses with n. The interpretation is that the sampling distribution of the sample mean degenerates to a point mass as n goes to infinity, since the variance goes to zero. A null variance means that the random variable at issue assumes one value only, so it is not quite random anymore, hence the term "degenerates". The point to which the variable "shrinks" to is actually μ. As we shall see later in the course, the property of an estimator to asymptotically shrink toward the quantity it is meant to estimate is a desirable and important one and is called consistency.

For the simulation part, I suggest the following simple code:

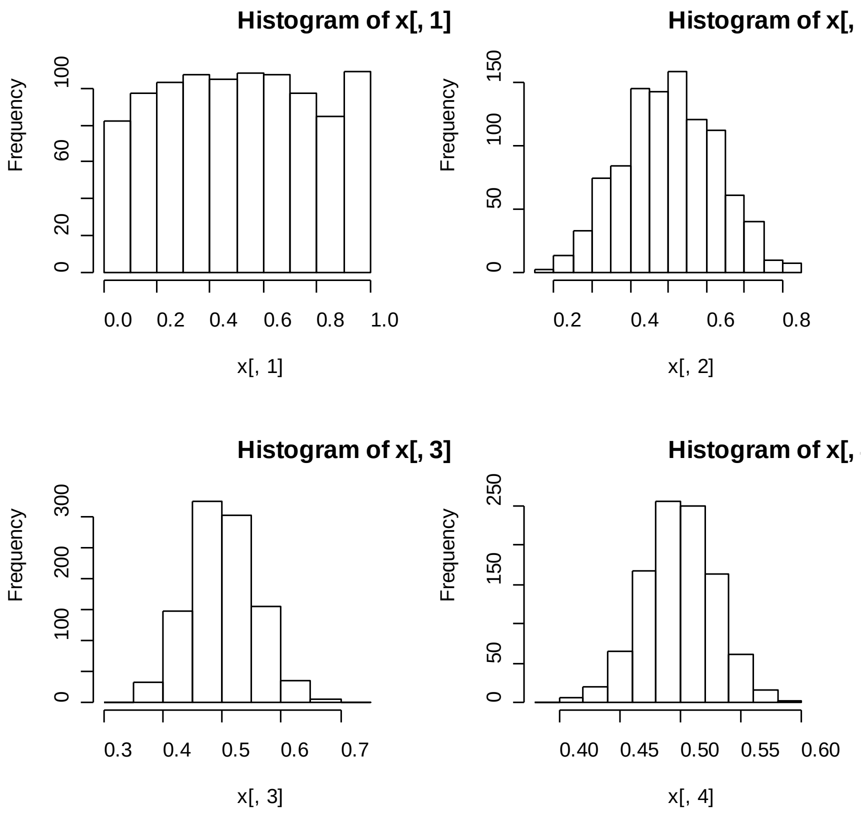
```
x <- matrix(0, 1000, 4)
size <- c(1, 5, 25, 100)
for (j in 1:4) {
for (i in 1:1000) {
x[i,j] <- mean(runif(size[j]))
}
}
```

where in the matrix x I store the simluated values of the sample mean columnwise, each column corresponding to a different sample size, as indicated in the text. Then we can instruct the following[1]

```
summary(x)[4,]
```

---

[1] Suggestion: play a little bit around with the function summary(), it applies to many objects (we used it for a matrix) and is very handy. Note that we just pulled a single row out of it.



```
"Mean :0.5054 " "Mean :0.5024 " "Mean :0.5011 " "Mean :0.4989 "
> (sd(x))^2
[1] 0.0804336467    0.0153208626    0.0030979086    0.0008837881
```

and we notice how the average is steady around 0.5 and the variance decreases, so supporting out theoretical findings.

Next let us look at the histograms of the simlulated sample average, for the different levels of n (figure above).

We notice how in this simple simulation two important theoretical results are at play: first the Central Limit Theorem (the distribution shapes from uniform to normal as n increases) and second the consistency of the sample mean: its distribution shrink more and more toward its mean (it might not look like this at first glance but pay attention to the range of the x-axis...).

Problem 4.1

$P\left(|x-\mu| \geq k\sigma\right)$ where we know that for the exponential distribution $\mu = \beta$ and $\sigma = \beta$. Hence:

$P\left(|x-\beta| \geq k\beta\right) = P\left(\beta - k\beta > x > \beta + k\beta\right) = P\left(\beta(1-k) > x > \beta(1+k)\right) = P(x < \beta(1-k)) + P(x > \beta(1+k))$

$= 0 + P(x > \beta(1+k))$ since $\beta, x > 0$ and $k > 1$

$= 1 - P(x < \beta(1+k)) = 1 - 1 + e^{-\frac{\beta(1+k)}{\beta}} = e^{-(1+k)}$ where we made use of the cdf for the exponential.

Substituting the values in Chebichev's inequality we find that the lower bound for the above probability is $\frac{1}{k^2}$ and it is easy to see that this number is always bigger than $e^{-(1+k)}$ for $k > 1$. In fact, for $k = 1$ we have $\mathrm{e}^{-2} < 1$

and as k increases the exponential decays faster than the power function, so that the first gets lower and lower than the second.

Problem 5.1

a) $E\left(S^2\right) = \sigma^2$.

Proof: $E\left(S^2\right) = \frac{1}{n-1}\sum E\left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum\left\{E(x_i^2) - 2E\left(x_i\overline{x}\right) + E\left(\overline{x}^2\right)\right\}$

now note that $E(x_i^2) = Var(x_i) + E^2(x_i) = \sigma^2 + \mu^2$

likewise: $E\left(\overline{x}^2\right) = \frac{\sigma^2}{n} + \mu^2$.

and: $E\left(x_i\overline{x}\right) = E\left(\sum_j \frac{x_j}{n}x_i\right) = E\left(\frac{x_i^2}{n} + \sum_j \frac{x_j}{n}x_i\chi\left\{j \neq i\right\}\right) = \frac{\sigma^2+\mu^2}{n} + \frac{n-1}{n}\mu^2$ (by linearity of expectation and independence).

Hence: $E\left(S^2\right) = \frac{n}{n-1}\left[\left(\sigma^2+\mu^2\right) - 2\left(\frac{\sigma^2+\mu^2}{n} + \frac{n-1}{n}\mu^2\right) + \left(\frac{\sigma^2}{n} + \mu^2\right)\right]$ (by identically distributed property)

A bit of simple algebra will take to : $\frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2 = E\left(S^2\right)$.

b) $S^2 = \frac{1}{n-1}\sum\left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum x_i^2 - \frac{2}{n-1}\sum x_i\overline{x} + \frac{n}{n-1}\left(\overline{x}^2\right)$

$= \frac{n}{n-1}\sum\frac{x_i^2}{n} - \frac{2n}{n-1}\left(\overline{x}^2\right) + \frac{n}{n-1}\left(\overline{x}^2\right) = \frac{n}{n-1}\sum\frac{x_i^2}{n} - \frac{n}{n-1}\left(\overline{x}^2\right)$.

Now let $c_n = d_n = \frac{n}{n-1}$. Both converge poinwise to 1 as $n \to \infty$. Furthermore, we can apply the Law of Large Numbers to state:

$\sum\frac{x_i^2}{n} \to_p E\left(x^2\right)$ and $\overline{x}^2 \to_p E\left(x\right)^2$. At this point, we can apply the sort of "linearity" property of the convergence in probability (stated in theorem 5.5) to conlude that $S^2 \to_p E\left(x^2\right) - E\left(x\right)^2 = \sigma^2$.

An observation is in order here. Many of you proved part b) considering the division by n instead of by n-1. With this caveat the calculations get a little easier. In this case you are fine, but this sort of manipulations are not risk free in general. Some of you attempted a justification, some did not bother at all. The right justification is that with that change we are not altering the rate of convergence of the random quantity, in that $\frac{1}{n-1}$ and $\frac{1}{n}$ go to zero with the same "speed". In other words, they are infinitesimal of the same order. Therefore, the random quantity will still have the same asymptotic behavior, as its rate of convergence has not changed. But if instead we had multiplied by $\sqrt{n}$ (instead of by $\frac{n-1}{n}$ as those of you did) we would have gotten a rate of convergence equal to $\frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$, which is not of the same order as $\frac{1}{n-1}$. It is actually slower, and somehow does not allow our random quantity to shrink fast enough about its mean (a constant fixed value), so that even at the limit it maintains its randomness, thus remaining a "proper" random variable. This would be the case of convergence in distribution, a much different situation than the convergence in probabilty. So the moral is whatch out with manipulations!

Problem 5.6

Let us start with a remark. Here, as many of you did, we do not have to assume that the height is normally distributed. the CLT does not allow us to go that far.

Rather, it allows us to approximate the true distribution of the height with the normal, but we have to keep this in mind, we are going to get approximative results from calculations involving the CLT, not exact ones. This said, we get:

$P(\overline{x} > 68) = P\left(\frac{\overline{x}-68}{2.6/100} > 0\right) \simeq P\left(z > 0\right) = 1/2$. Where z is a variable with standard normal distribution and the approximate equality follows by the CLT.