hw_lect15_1

Examine hw_lect7_1. In our new language, what you did there is to find the sampling distribution of the sample minimum (for samples of size n=3).
a) Revise the posted solution to find the sampling distribution of the sample mean for n=3.
b) Show that the mean (expected value) of that distribution is pi.

a)

	$\bar{x}$	pr
000	0	$(1-\pi)^3$
001	1/3	$(1-\pi)^2\pi$
010	1/3	
100	1/3	
011	2/3	$(1-\pi)\pi^2$
101	2/3	
110	2/3	
111	1	π^3

$\bar{x}$	0	1/3	2/3	1
$p(\bar{x})$	$(1-\pi)^3$	$3\pi(1-\pi)^2$	$3(1-\pi)\pi^2$	π^3

FYI: This $p(\bar{x})$ is clearly not Normal! And that may seem to contradict the CLT. The reason is that CLT applies to continuous dists, while here we are dealing with a discrete dist. (Bernoulli).

b) $E[\bar{x}] = \sum_{\bar{x}} \bar{x}\, p(\bar{x}) = 0(1-\pi)^3 + \frac{1}{3} 3\pi(1-\pi)^2 + \frac{2}{3} 3\pi^2(1-\pi) + 1\pi^3$

$= \pi(1-\pi)^2 + 2\pi^2(1-\pi) + \pi^3$

$= \pi(1-\pi)[1-\pi+2\pi] + \pi^3 = \pi(1-\pi)(1+\pi) + \pi^3$

$= \pi(1-\pi^2) + \pi^3 = \pi - \pi^3 + \pi^3 = \pi.$

Note: Here, the population distr. is Bernoulli(π), and we already know from Ch.2 that for Binomial(π), we have $\mu_x = \pi$. So, the calculation in this hw confirms the general result $E[\bar{x}] = \mu_x$, even when the distr. is not normal.

I did not ask you to find $V[\bar{x}]$, but if you do you'll find:

$$V[\bar{x}] = \sum_{\bar{x}} (\bar{x}-\mu_x)^2 p(\bar{x}) = (0-\pi)^2(1-\pi)^3 + \left(\tfrac{1}{3}-\pi\right)^2 3\pi(1-\pi)^2 + \left(\tfrac{2}{3}-\pi\right)^2 3(1-\pi)\pi^2 + (1-\pi)^2\pi^3$$

$= \cdots$ lots of algebra $\cdots = \dfrac{\pi(1-\pi)}{3} \implies V[\bar{x}] = \dfrac{\pi(1-\pi)}{3}$

This confirms $V[\bar{x}] = \dfrac{\sigma_x^2}{n}$, because for Bernoulli($\pi$), we already know that $\sigma_x^2 = \pi(1-\pi)$.

hw_lect15_2 (By R)
a) write R code to produce the sampling distribution of the sample maximum, for samples of size 50 taken from a standard Normal. Use 5000 trials,
b) Repeat for the sample minimum.

Turn-in your code, and the resulting 2 histograms.
FYI, these distributions arise naturally when one tries to model extreme events, e.g. the biggest storms, the strongest earthquakes,the brightest stars, the smallest forms of life, etc.

```
ntrial = 5000
xmax = xmin = numeric(ntrial)
for(trial in 1:ntrial){
x = rnorm(50,0,1)
xmax[trial] = max(x)
xmin[trial] = min(x)
}
hist(xmax)
hist(xmin)
```

hw_lect15_3 (By R)
a) write R code to take 5000 samples of size n=100 from an exponential distr. with parameter lambda=2, and make a qq-plot of the 5000 means. Recall that if the qq-plot is a straight line, then the histogram of the sample means is Normal. This will show that the sampling distr. of sample means is Normal, even when the pop. is not!

b) using the qq-plot, estimate the mean and std. dev. of the sampling dist. of sample means. Are they consistent with what you would expect from our formulas for the mean and standard deviation of the sampling distribution? show work. $\leftarrow$ $\mu_{\bar{x}} = \mu_x$, $\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$

```
# a)
ntrial = 5000
xbar = numeric(ntrial)
for(trial in 1:ntrial){
x = rexp(100,2)
xbar[trial] = mean(x)
}
qqnorm(xbar)
abline(0.5, 0.5/sqrt(100))
```

b)
According to these formulae, the y-intercept of the qq-plot should be mu_x, which for our exponential dist is 1/2=0.5. The slope of the qq-plot should be sigma_x/sqrt(n). Recall that for exponential we have mu_x = 1/lambda = 1/2. From the abline above, we see that these are correct intercept and slope for the qq-plot.

hw_lect15_4

A sample of size 36 from a Normal pop. yields the observed values xbar= 3.5 and s=1.

a) Under the assumption that mu_x = 2.5, and sigma_x = 2, what's the prob of a sample mean larger than the one observed?

b) Under the assumption that mu_x = 2.5, and sigma_x = 2, what's the prob of a sample mean smaller than the one observed?

c) Under the assumption that mu_x = 3.5, and sigma_x = 2, what's the prob of a sample mean larger than the one observed?

d) Under the assumption that mu_x = 3.5, and sigma_x = 2, what's the prob of a sample mean smaller than the one observed?

e) Now, suppose we know that sigma_ = 2, but we don't know mu_x. What is the observed 95% Confidence Interval for mu_x. Interpret it, in TWO ways. postpone e)

Table I.

a) $P(\bar{x} > \bar{x}_{obs}) = P(z > z_{obs}) = P\left(z > \frac{3.5-2.5}{2/\sqrt{36}}\right) = P(z > 3) = 1 - 0.9987 = .0013$

If $\mu_x = 2.5$, and $\sigma_x = 2$, it is unlikely to get a sample mean larger than that observed.

b) $P(\bar{x} < \bar{x}_{obs}) = P(z < z_{obs}) = P(z < 3) = 0.9987$

If $\mu_x = 2.5$, and $\sigma_x = 2$, it is likely to get a sample mean smaller than that observed.

c) $P(\bar{x} > \bar{x}_{obs}) = P(z > z_{obs}) = P\left(z > \frac{3.5-3.5}{2/\sqrt{36}}\right) = P(z > 0) = 0.5$

If $\mu_x = 3.5$, and $\sigma_x = 2$, it is equally likely to get a sample mean larger than or smaller than that observed.

d) $P(\bar{x} < \bar{x}_{obs}) = P(z < z_{obs}) = \text{prob}(z < 0) = 0.5$

same as in part c.

hw_lect16-1

e) $\bar{x}_{obs} \pm 1.96 \frac{\sigma_x}{\sqrt{n}} = 3 \pm 1.96 \frac{2}{\sqrt{36}} = 3 \pm 0.65 = (2.3, 3.6)$

- We are 95% confident that μ_x is between 2.3 and 3.6.
- There is 95% prob. that a random CI will cover the true mean μ_x.

hw-lect16-2

A sample of size 25 yields $\bar{x}_{obs} = 3$, $s_{obs} = 1.5$. Suppose we know $\sigma_x = 1$. Then, as shown in the example, we can be 95% confident that μ_x is in the interval $(2.6, 3.4)$.

a) Based on this information, can you find $pr(2.6 < \bar{x} < 3.4)$? If yes, find it. If not, why not?

No. Computing probs involves standardizing, which itself involves μ_x, σ_x. So, if we don't know the pop. distr., then we cannot get probs.

b) Show $pr(2.6 < \bar{x} < 3.4) = pr(z_{obs} - 1.96 < z < z_{obs} + 1.96)$, where $z_{obs} = (\bar{x}_{obs} - \mu_x)/(\sigma_x/\sqrt{n})$. Hint: $2.6 = \bar{x}_{obs} - 1.96 \frac{\sigma_x}{\sqrt{n}}$.

$$pr(2.6 < \bar{x} < 3.4) = pr\left(\frac{2.6 - \mu_x}{\sigma_x/\sqrt{n}} < \frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} < \frac{3.4 - \mu_x}{\sigma_x/\sqrt{n}} \right)$$

$$= pr\left(\frac{\bar{x}_{obs} - 1.96\frac{\sigma_x}{\sqrt{n}} - \mu_x}{\sigma_x/\sqrt{n}} < z < \frac{\bar{x}_{obs} + 1.96\frac{\sigma_x}{\sqrt{n}} - \mu_x}{\sigma_x/\sqrt{n}} \right)$$

$$= pr(z_{obs} - 1.96 < z < z_{obs} + 1.96)$$

Moral: The prob of a random $\bar{x}$ falling in the obs. 95% CI is NOT 95%!

c) Now, forget all observed information, and suppose we know $\mu_x = 2$, $\sigma_x = 1$. What is the numerical value of $pr(2.6 < \bar{x} < 3.4)$ when $n = 25$?

$$pr(2.6 < \bar{x} < 3.4) = pr\left(\frac{2.6 - 2}{1/\sqrt{25}} < \frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} < \frac{3.4 - 2}{1/\sqrt{25}} \right) = pr(3 < z < 7)$$

$$= pr(z < 7) - pr(z < 3) = 1 - .9987 = \boxed{0.0013} \neq 0.95$$

Moral: the $pr(2.6 < \bar{x} < 3.4)$ is 2 different numbers in parts b & c. In part b, the numbers 2.6 and 3.4 are based on observed data. In part c, they are not. That changes the answers!

e) How about $\mu_x = 2.8$ and $\sigma_x = 0.608$; find $prob(2.6 < \bar{x} < 3.4)$?

$$pr(2.6 < \bar{x} < 3.4) = pr\left(\frac{2.6 - 2.8}{0.608/\sqrt{25}} < \frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} < \frac{3.4 - 2.8}{.608/\sqrt{25}} \right) = pr(-1.645 < z < 4.9)$$

$$= pr(z < 4.9) - pr(z < -1.645) = 1 - .05 = \boxed{0.95}$$

Moral: If you insist on having $pr(2.6 < \bar{x} < 3.4) = 0.95$, then a possible value of μ_x and σ_x are 2.8 and 0.608. But those are not unique.

hw-lect 16 – 3

One can build C.I for any pop. param (not just the mean μ_x).
For example, if the population is Unif(0, b), we can build a CI for the $\underline{b}$ param. Do it!
zero

Assume $\bar{x}$ is (approximately) Normal: Recall that μ_x and σ_x of Unif(a,b) are $\frac{1}{2}(a+b)$, $\frac{(b-a)}{\sqrt{12}}$, respectively.

Hints: what is the quantity that has a normal dist?
what is " " " " " standard norm. dist?
Use it in a self-evident fact, and solve for $\underline{b}$.

$$\bar{x} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right) = N\left(\frac{b+a}{2}, \frac{b-a}{\sqrt{12n}}\right) = N\left(\frac{b}{2}, \frac{b}{\sqrt{12n}}\right)$$

$$\therefore z = \frac{\bar{x} - \frac{b}{2}}{\frac{b}{\sqrt{12n}}} \sim N(0,1).$$

$$pr(-z^* < z < z^*) = \text{Conf. level}$$

$$-z^* < \frac{\bar{x} - \frac{b}{2}}{\frac{b}{\sqrt{12n}}} < z^*$$

$$-z^* < \sqrt{12n}\left(\frac{\bar{x}}{b} - \frac{1}{2}\right) < z^*$$

$$-\frac{z^*}{\sqrt{12n}} < \frac{\bar{x}}{b} - \frac{1}{2} < z^*/\sqrt{12n}$$

$$\frac{1}{2} - \frac{z^*}{\sqrt{12n}} < \frac{\bar{x}}{b} < \frac{1}{2} + \frac{z^*}{\sqrt{12n}}$$

$$\frac{1}{\frac{1}{2} + \frac{z^*}{\sqrt{12n}}} < \frac{b}{\bar{x}} < \frac{1}{\frac{1}{2} - \frac{z^*}{\sqrt{12n}}}$$

$$\frac{\bar{x}}{\frac{1}{2} + \frac{z^*}{\sqrt{12n}}} < b < \frac{\bar{x}}{\frac{1}{2} - \frac{z^*}{\sqrt{12n}}} \Rightarrow$$

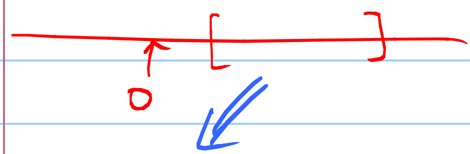
$$\boxed{\text{C.I. for } b: \quad \frac{2\bar{x}}{1 \pm \frac{z^*}{\sqrt{3n}}}}$$

Think about it; This is super-cool! If you believe your data come from a Unif(0, b), you find the sample mean ($\bar{x}$) of your data, and then this formula tells you something about what $\underline{b}$ can be!

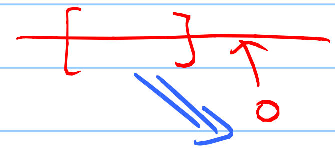
hw_optional

Suppose $\sigma_x/\sqrt{n} = 1$. Find the possible range of $\bar{x}_{obs}$ values such that the 95% obs. CI for μ_x allows for μ_x to be zero.

$$\bar{x}_{obs} \pm 1.96 \; 1 = [\bar{x}_{obs} - 1.96, \; \bar{x}_{obs} + 1.96]$$



or



$$\bar{x}_{obs} - 1.96 < 0 \Rightarrow \underline{-1.96 < \bar{x}_{obs} < 1.96} \Leftarrow \bar{x}_{obs} + 1.96 > 0$$

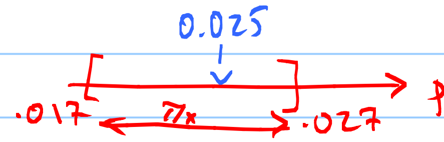
hw_lect17_1

A sample of 2000 aluminum screws used in the assembly of electronic components was examined, and it was found that 44 of these screws stripped out during the assembly process. Does it appear that the true percentage of defective screws is (or is not) 2.5%? Explain your reasoning and the conclusion that follows from it. You may use the "simple formula" appropriately revised. Use 90% confidence level

The 90% CI for π_x:

$$p \pm 1.645\sqrt{\frac{p(1-p)}{n}} \qquad \text{where } p = \frac{44}{2000} = .022 .$$

$$.022 \pm 1.645\sqrt{\frac{.022(.978)}{2000}} = 0.022 \pm .0054 = (.017, .027)$$



Because 2.5% falls in the CI, all we can say is that the true prop. of defective screws <u>may</u> be 2.5%. We cannot know if it really is 2.5%; all we know is that the true prop. is somewhere between 0.017 and 0.027 (with 90% conf.) So, do <u>NOT</u> say that there is evidence that the true prop. is 2.5%.

By contrast if the CI did not include 2.5%, then we could say that the true prop. is not 2.5% (with 90% conf.)

hw-lect17-2

There are several ways of proving $E[p] = \pi$, $V[p] = \sqrt{\frac{\pi(1-\pi)}{n}}$, (dropping the subscript x, just for convenience). One way is to use a result which we have already derived, ie. $E[\bar{x}] = \mu_x$, $V[\bar{x}] = \frac{\sigma_x^2}{n}$. This result holds even if the x_i are 0 or 1. So, first,

a) Consider a sample of size n from a Bernoulli distribution, ie. n zeros and 1's, and show that the sample mean ($\bar{x}$) is equal to the sample proportion of 1's. Hint: if a sample of size n has n_0 0's and n_1 1's, then the sample prop. p is $\frac{n_1}{n}$.

$$\bar{x} = \frac{1}{n}\sum_i^n x_i = \frac{1}{n}\Big[\overbrace{0+0+\cdots+0}^{n_0} + \overbrace{1+1+\cdots 1}^{n_1}\Big] = \frac{n_1}{n} = p$$

So, at this point, it follows that $E[p] = E[\bar{x}]$, $V[p] = V[\bar{x}]$

But we already know that $E[\bar{x}] = \mu_x$ and $V[\bar{x}] = \sigma_x^2/n$, where μ_x and σ_x^2 are the dist. mean and dist. var. of variable taking only 0, 1 values, ie. a Bernoulli random variable. So,

b) For $x \sim$ Bernoulli(π), find μ_x and σ_x^2 starting from the definition of $E[x]$ and $V[x]$ from ch. 2.

$$\mu_x = E[x] = \sum_{x=0}^{1} x\, p(x) = 0\, p(0) + 1\, p(1) = \pi$$

$$\sigma_x^2 = V[x] = \sum_{x=0}^{1} (x-\mu_x)^2 p(x) = (0-\pi)^2 p(0) + (1-\pi)^2 p(1)$$

$$\pi^2(1-\pi) + (1-\pi)^2\pi = \pi(1-\pi)$$

Moral: { When you are done, you will have proven $E[p] = \pi$, $V[p] = \frac{\pi(1-\pi)}{n}$ using equations that we had proven before, ie. $E[\bar{x}] = \mu_x$, $V[\bar{x}] = \frac{\sigma_x^2}{n}$.

$$\begin{array}{ccc} E[\bar{x}] & = & \mu_x \\ a \rightarrow \| & & \| \leftarrow b \\ E[p] & & \pi \end{array} \qquad \begin{array}{ccc} V[\bar{x}] & = & \sigma_x^2/n \\ a \rightarrow \| & & \| \leftarrow b \\ V[p] & & \frac{\pi(1-\pi)}{n} \end{array}$$