Math 262

Section 4.5 Day 34

1. Let random variable X have one of the following distributions. For what distribution of iid random variables $Y_1, Y_2, \ldots, Y_n$ is it the case that $X = Y_1 + Y_2 + \cdots + Y_n$?

   (a) $X \sim \text{Bin}(n, p)$

   $Y_i \sim \text{Bernoulli}(p)$

   X is approx. normal when n is big (and p is not too close to 0 or 1)

   (b) $X \sim \text{Gamma}(\alpha = n, \beta)$

   $Y_i \sim \text{Exp}\left(\lambda = \frac{1}{\beta}\right)$

   X is approx. normal when α is large

   (c) $X \sim \text{Poisson}(\lambda = n)$

   $Y_i \sim \text{Poisson}(1)$

   X is approx. normal when λ is large

   (d) $X \sim \text{NegBin}(r = n, p)$

   $Y_i \sim \text{Geometric}(p)$

   X is approx. normal when r is large

   CENTRAL LIMIT THEOREM

2. Customers at a popular restaurant are waiting to be served. Waiting times are independent and exponentially distributed with mean $1/\lambda = 10$ minutes.

   (a) What is the probability that the average wait time of the 50 customers is less than 12 minutes?

   Total of 50 waiting times: $T_{50} \sim \text{Gamma}(\alpha = 50, \beta = 10)$

   Var: $\alpha\beta^2$

   σ: $10\sqrt{50}$

   Average waiting time: $\bar{X}_{50} = \frac{T_{50}}{50}$

   $P(\bar{X}_{50} < 12) = P\left(\frac{T_{50}}{50} < 12\right) = P(T_{50} < 600) \approx 0.916$

   (b) Use a normal distribution to approximate the probability that the average wait time of 50 customers is less than 12 minutes. What limit theorem justifies this?

   Since $\alpha = 50$ is "large", $\bar{X}_{50}$ is approx. normal $Z \sim N\left(\mu = 10, \sigma = \frac{10}{\sqrt{50}}\right)$

   $P(\bar{X}_{50} < 12) \approx P(Z < 12) \approx 0.921$

3. Suppose you flip a fair coin *lots* of times. What does the Law of Large Numbers say about the numbers of heads and tails you will observe?

The numbers of heads and tails <u>must</u> approach 50% of the number of coin flips.

4. Let $X_1, X_2, \ldots, X_n$ be iid random variables with an Exp($\lambda = 2$) distribution. Let $\mu = E(X_i)$.

so $\mu = \frac{1}{2}$

   (a) What is the distribution of T_n? What is the value of μ?

$T_n = X_1 + X_2 + \cdots + X_n$

$T_n \sim \text{Gamma}\left(\alpha = n,\ \beta = \frac{1}{2}\right)$

$\bar{X}_n = \frac{T_n}{n}$

   (b) In R or *Mathematica*, write a function that computes $P\left(\left|\frac{T_n}{n} - \mu\right| \geq \epsilon\right)$ for any given parameter values n and ϵ.

Probability that $\bar{X}_n = \frac{T_n}{n}$ is at least ε away from the true mean μ

$= 1 - P\left(\mu - \varepsilon < \frac{T_n}{n} < \mu + \varepsilon\right)$

$= 1 - P\left(n\mu - n\varepsilon < T_n < n\mu + n\varepsilon\right)$

   (c) Make a plot of $P\left(\left|\frac{T_n}{n} - \mu\right| \geq 0.01\right)$ for values of n between 1 and 10,000. What limit theorem does this plot illustrate?

See Mathematica file!

<u>Central Limit Theorem</u>

   (d) What is the smallest n such that $P\left(\left|\frac{T_n}{n} - \mu\right| \geq 0.01\right) < 0.01$?

5. Suppose that a fair coin is tossed 1000 times. If the first 100 tosses all result in heads, what proportion of heads would you expect on the remaining 900 tosses? Interpret the statement "The law of large numbers swamps, but it does not compensate."

1. Let random variable X have one of the following distributions. For what distribution of iid random variables $Y_1, Y_2, \dots, Y_n$ is it the case that $X = Y_1 + Y_2 + \cdots + Y_n$?

(a) $X \sim \text{Bin}(n, p)$ $Y_i \sim \text{Bernoulli}(p)$

X is approx. normal when $np \geq 10$ and $n(1-p) \geq 10$.

(b) $X \sim \text{Gamma}(\alpha = n, \beta)$ $Y_i \sim \text{Exp}\left(\lambda = \frac{1}{\beta}\right)$

X is approx. normal when α is large.

(c) $X \sim \text{Poisson}(\lambda = n)$ $Y_i \sim \text{Poisson}(1)$

X is approx. normal when λ is large.

(d) $X \sim \text{NegBin}(r = n, p)$ $Y_i \sim \text{Geom}(p)$

X is approx. normal when r is large.

2. Customers at a popular restaurant are waiting to be served. Waiting times are independent and exponentially distributed with mean $1/\lambda = 10$ minutes.

(a) What is the probability that the average wait time of the 50 customers is less than 12 minutes?

T_{50} is $\text{Gamma}(\alpha = 50, \beta = 10)$.

$\bar{X}_{50} = \frac{T_{50}}{50}$ $P(\bar{X}_{50} < 12) = P\left(\frac{T_{50}}{50} < 12\right) = P(T_{50} < 600) \approx 0.916$

$\bar{X}_{50}$ is $\text{Gamma}(\alpha = 50, \beta = \frac{1}{5})$. Why? mgfs!

R: pgamma(600, 50, $\frac{1}{10}$)

(b) Use a normal distribution to approximate the probability that the average wait time of 50 customers is less than 12 minutes. What limit theorem justifies this?

$\bar{X}_{50}$ is approx $N\left(10, \frac{10}{\sqrt{50}}\right)$, so $P(\bar{X}_n < 12) \approx P(Z < 12) \approx 0.921$

$Z \sim N\left(10, \frac{10}{\sqrt{50}}\right)$

R: pnorm(12, 10, 1.414)

3. Suppose you flip a fair coin *lots* of times. What does the Law of Large Numbers say about the numbers of heads and tails you will observe?

These numbers must approach 50% of the number of flips.

4. Let $X_1, X_2, \dots, X_n$ be iid random variables with an Exp($\lambda = 2$) distribution. Let $\mu = E(X_i)$.

(a) What is the distribution of T_n? What is the value of μ?

$T_n \sim \text{Gamma}(\alpha = n, \beta = \frac{1}{2})$ $\qquad \mu = E(X_i) = \frac{1}{2}$

(b) In **R** or *Mathematica,* write a function that computes $P\left(\left|\frac{T_n}{n} - \mu\right| \geq \epsilon\right)$ for any given parameter values n and ϵ.

First: $P\left(\left|\frac{T_n}{n} - \mu\right| \geq \varepsilon\right) = 1 - P\left(\left|\frac{T_n}{n} - \frac{1}{2}\right| < \varepsilon\right) = 1 - P\left(-\varepsilon < \frac{T_n}{n} - \frac{1}{2} < \varepsilon\right)$

$$= 1 - P\left(\frac{n}{2} - n\varepsilon < T_n < \frac{n}{2} + n\varepsilon\right)$$

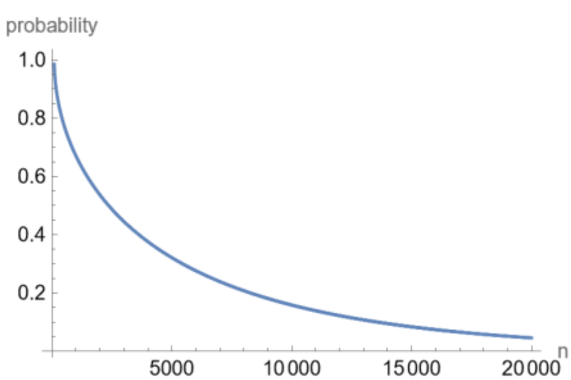
R:

```
wlln <- function(n, eps){
  1 - (pgamma(n/2 + n*eps, n, 2) - pgamma(n/2 - n*eps, n, 2))
}
```

Mathematica:

```
wlln[n_, ϵ_] := 1 - Probability[n/2 - n*ϵ < Tn < n/2 + n*ϵ, Tn ≈ GammaDistribution[n, 1/2]]
```

(c) Make a plot of $P\left(\left|\frac{T_n}{n} - \mu\right| \geq 0.01\right)$ for values of n between 1 and 10,000. What limit theorem does this plot illustrate?



$P\left(\left|\frac{T_n}{n} - \mu\right| \geq \frac{1}{100}\right)$ converges to zero, illustrating the weak law of large numbers.

(d) What is the smallest n such that $P\left(\left|\frac{T_n}{n} - \mu\right| \geq 0.01\right) < 0.01$?

By trial and error, we find $n = 16{,}589$.

5. Suppose that a fair coin is tossed 1000 times. If the first 100 tosses result in heads, what proportion of heads would you expect on the remaining 900 tosses? Interpret the statement "The law of large numbers swamps, but it does not compensate."

We expect about 450 heads in the remaining 900 tosses.

Any given observations do not change the probabilities for later tosses. However, even a very unusual sequence of heads will be insignificant in the long run as the proportion of heads will converge to $\frac{1}{2}$.