Stat 311: HW 6, Chapter 8, Solutions

Fritz Scholz

Ch. 8, Problem 2. The law of large numbers states that the empirical distribution $\hat{P}_n$ converges to the true (sampled) distribution P. Since the latter is apparently not normal, $\hat{P}_n$ will certainly not look normal for very large samples. The CLT only comes into play with respect to the approximate distribution of random variable $\hat{P}_n(A)$, namely

$$\hat{P}_n(A) \approx \mathcal{N}(P(A), P(A)(1-P(A))/n)$$

but this statement examines how $\hat{P}_n(A)$ varies around its expected value $P(A)$ and that variation is indeed normal. However, the kernel density estimate based on a large sample $X_1,\ldots,X_n$ will estimate the sampled distribution which was not normal (bimodal).

Ch. 8, Problem 4. Let X_i denote the usage time delivered by the i^{th} pair of batteries. $EX_i = 5$ h and $\sigma(X_i) = 0.5$ h. If $Y = X_1 + \ldots + X_{20}$ we are interested in $P(Y \geq 105)$. $EY = 20 \times 5 = 100$ h and $\sigma(Y) = 0.5 \times \sqrt{20} = \sqrt{5} = 2.236068$.

$$\begin{aligned} P(Y \geq 105) = P\left(\frac{Y-100}{\sqrt{5}} \geq \frac{105-100}{\sqrt{5}}\right) &\approx P(Z \geq 5/\sqrt{5}) = P(Z \geq \sqrt{5}) \\ &= P(Z \geq 2.236068) = 1 - \texttt{pnorm}(2.236068) = 0.01267366 \end{aligned}$$

i.e., his chances are quite slim $(.0123)$ that he can get at least 105 hours out of his 20 2-packs.

Ch. 8, Problem 5. Let X_i denote the number on the i^{th} drawn ticket. Using `x <- c(1,1,1,1,2,5,5,10,10,10)` and `n <- length(x)` we find $n = 10$, $EX_i = \texttt{mean(x)} = 4.6$ and $\text{var} X_i = \sigma^2(X_i) = \texttt{var(x)} * (\texttt{n}-1)/\texttt{n} = 14.64$. For $Y = X_1 + \ldots + X_{40}$ we get $EY = 40 \cdot 4.6 = 184$ and $\text{var} Y = 40 \cdot 14.64 = 585.6$ and thus $\sigma(Y) = \texttt{se} = \texttt{sqrt(585.6)}$ $= 24.19917$.

```
(a)
> x <- c(1,1,1,1,2,5,5,10,10,10)

urn.model <- function (pop=x,n=40)
{
y.vec <- sample(pop,n,replace=T)
sum(y.vec)
}

urn.model.sim <- function (Nsim=1000,n=40)
{
out <- numeric(Nsim)
for(i in 1:Nsim){
out[i] <- urn.model(x,n)
}
mean(out > 170.5 & out < 199.5)
}

> urn.model.sim(100000)
[1] 0.44573
> urn.model.sim(100000)
[1] 0.44609
> urn.model.sim(100000)
[1] 0.44767
```

```
> urn.model.sim(100000)
[1] 0.44626
> urn.model.sim(100000)
[1] 0.44478

> mean(c(0.44573,0.44609,0.44767,0.44626,0.44478))
[1] 0.446106
```

Her reasoning is justified by the law of large numbers because the proportion of values falling within $(170.5, 199.5)$ is an average of i.i.d. Bernoulli random variables with success probability $p = P(170.5 < Y < 199.5)$. By taking a large number of simulated Bernoulli trials we should get very close to the true p. The above simulations appear to show values that closely scatter around 0.445.

```
(b)
> se <- sqrt(585.6)
> pnorm(199.5,mean=184,sd=se)-pnorm(170.5,mean=184,sd=se)
[1] 0.4506155
```

The approximation is based on the CLT for $Y = X_1 + \ldots + X_{40}$ with mean and variance given previously, i.e., $Y \approx \mathcal{N}(\mu = 184, \sigma^2 = 585.6)$.

$$P(170.5 < Y < 199.5) = \texttt{pnorm}(199.5, \texttt{mean} = 184, \texttt{sd} = \texttt{se}) - \texttt{pnorm}(170.5, \texttt{mean} = 184, \texttt{sd} = \texttt{se})$$

(c) The approximation is not too far off from our simulated results, but it is well outside the scatter of the various simulations. The normal approximation cannot be improved for this sample size $n = 40$, while we can get a better simulated answer by increasing the number of simulations.

Ch. 8, Problem 6. (a) We want to evaluate (approximate) $P(V > 0)$ where $V = X_1 + \ldots + X_{400}$. $EV = 400 \cdot 0.01 = 4$ and $\text{var}\, V = 400 \cdot 0.01 = 4$. Thus

$$P(V > 0) = 1 - P(V \leq 0) \approx 1 - \texttt{pnorm}(0, 4, 2) = 0.9772499$$

(b) We want to evaluate (approximate) $P(W > 0)$ where $W = Y_1 + \ldots + Y_{400}$. $EW = 400 \cdot 0 = 0$ and $\text{var}\, V = 400 \cdot 0.25 = 100$. Thus

$$P(W > 0) = 1 - P(W \leq 0) \approx 1 - \texttt{pnorm}(0, 0, 10) = 0.5$$

(c) We want to evaluate (approximate) $P(V \geq 20)$. Thus

$$P(V \geq 20) = 1 - P(V < 20) \approx 1 - \texttt{pnorm}(20, 4, 2) = 6.661338\text{e} - 16$$

(d) We want to evaluate (approximate) $P(W \geq 20)$. Thus

$$P(W \geq 20) = 1 - P(W < 20) \approx 1 - \texttt{pnorm}(20, 0, 10) = 0.02275013$$

(e) We evaluate $P(W - V > 0)$. Since $W \approx \mathcal{N}(0, 100)$ and $V \approx \mathcal{N}(4, 4) \Longrightarrow W - V \approx \mathcal{N}(0 - 4, 100 + 4)$ and thus

$$P(W - V > 0) = 1 - P(W - V \leq 0) \approx 1 - \texttt{pnorm}(0, -4, \texttt{sqrt}(104)) = 0.3474433$$