1. Suppose that you take a random sample of 42 laptops. The startup time for the laptops follows some distribution with a mean of 10 seconds and a standard deviation of 4 seconds. If the average startup time of your sample is greater than 11 seconds, you must return the entire sample. What is the probability that the laptops must be returned?

Since the sample size is 42, by CLT:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[D]{} N(0, \sigma^2)$$

Therefore:

$$\bar{X}_n \text{ approximately follows: } N\left(\mu, \frac{\sigma^2}{n}\right)$$

That is: $N\left(10, \frac{4^2}{42}\right)$. $P(\bar{X}_n > 11) = P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{11-10}{4/\sqrt{42}}\right)$

in which, $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

\# R: $1 - \text{Pnorm}\left(\frac{1}{4/\sqrt{42}}\right) \approx 0.0526$

2. CB 5.3

5.3 Let $X_1, \ldots, X_n$ be iid random variables with continuous cdf F_X, and suppose $\mathrm{E}X_i = \mu$. Define the random variables $Y_1, \ldots, Y_n$ by

$$Y_i = \begin{cases} 1 & \text{if } X_i > \mu \\ 0 & \text{if } X_i \le \mu. \end{cases}$$

Find the distribution of $\sum_{i=1}^n Y_i$.

Y_i is i.i.d. Bernoulli distribution with $P(Y_i = 1) = 1 - F_X(\mu)$

Therefore: $\sum_{i=1}^n Y_i$ is binomial distribution,

$$P(Y_i = k) = \binom{n}{k} [F_X(\mu)]^{n-k} [1 - F_X(\mu)]^k$$

3. CB 5.4

5.4 A generalization of iid random variables is *exchangeable* random variables, an idea due to deFinetti (1972). A discussion of exchangeability can also be found in Feller (1971). The random variables $X_1, \ldots, X_n$ are *exchangeable* if any permutation of any subset of them of size k ($k \le n$) has the same distribution. In this exercise we will see an example of random variables that are exchangeable but not iid. Let $X_i|P \sim$ iid Bernoulli(P), $i = 1, \ldots, n$, and let $P \sim$ uniform$(0,1)$.

(a) Show that the marginal distribution of any k of the Xs is the same as

$$P(X_1 = x_1, \ldots, X_k = x_k) = \int_0^1 p^t (1-p)^{k-t} dp = \frac{t!(k-t)!}{(k+1)!},$$

where $t = \sum_{i=1}^k x_i$. Hence, the Xs are exchangeable.

(b) Show that, marginally,

$$P(X_1 = x_1, \ldots, X_n = x_n) \neq \Pi_{i=1}^n P(X_i = x_i),$$

so the distribution of the Xs is exchangeable but not iid.

(a) By conditional independence:

$$P(X_1 = x, \cdots, X_k = x_k | p) = \prod_{i=1}^k P(X_i = x_i | p) = p^t (1-p)^{k-t}$$

Joint: $f(x_1, x_2 \cdots x_k, p) = f(x_1, x_2 \cdots x_k | p) \cdot f(p) = p^t (1-p)^{k-t}$

Marginal: $f(x_1, x_2 \cdots x_k) = \int_0^1 \underbrace{p^t (1-p)^{k-t}}_{\text{It's a Beta Kernel: Beta}(t+1,\, k-t+1)} dp = B(t+1, k-t+1)$

$$= \frac{\Gamma(t+1)\Gamma(k-t+1)}{\Gamma(k+2)} = \frac{t!\,(k-t)!}{(k+1)!}$$

(b)

Marginally : $P(X_i = x_i) = \int_0^1 f(X_i = x_i \mid p) f(p)\, dp$

$= \int_0^1 p^{x_i} (1-p)^{1-x_i}\, dp$

$P(X_i = 1) = \int_0^1 p\, dp = \frac{1}{2}$ $\quad P(X_i = 0) = \int_0^1 (1-p)\, dp = \frac{1}{2}$

$\prod_{i=1}^{k} P(X_i = x_i) = \frac{1}{2^k} \neq \text{Beta}(t+1, k-t+1)$

4. CB 5.13

5.13 Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$. Find a function of S^2, the sample variance, say $g(S^2)$, that satisfies $\mathrm{E}g(S^2) = \sigma$. (*Hint*: Try $g(S^2) = c\sqrt{S^2}$, where c is a constant.)

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\sum_{i=1}^{n} x_i^2 - n\bar{X}_n^2\right) \cdot \frac{1}{n-1}$

WLOG, assume that $\mu = 0$.

Since x_i are i.i.d. distributed : $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, let $t = \frac{(n-1)S^2}{\sigma^2}$

$E(\sqrt{S^2}) = E\left(\sqrt{\frac{\sigma^2 t}{n-1}}\right) = \sqrt{\frac{\sigma^2}{n-1}}\, E(\sqrt{t})$

$E(t) = \int_0^\infty \sqrt{t} \cdot \frac{(\frac{1}{2})^{k/2}}{\Gamma(k/2)} \cdot t^{\frac{k}{2}-1} e^{-\frac{t}{2}}\, dt$ $(k = n-1)$

$= \frac{(\frac{1}{2})^{k/2}}{\Gamma(k/2)} \cdot \int_0^\infty \underbrace{t^{\frac{k}{2}-\frac{1}{2}} e^{-\frac{t}{2}}}_{\text{Gamma Kenel, Gamma}(\frac{k+1}{2}, \frac{1}{2})}\, dt$

$= \frac{(\frac{1}{2})^{k/2}}{\Gamma(k/2)} \cdot \frac{\Gamma(\frac{k+1}{2})}{(\frac{1}{2})^{\frac{k+1}{2}}} = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(k/2)} \cdot \sqrt{2}$, $k = n-1$

$$E(\sqrt{S^2}) = \sqrt{\frac{\sigma^2}{n-1}}\, E(\sqrt{t}) = \sigma \cdot \sqrt{\frac{1}{n-1}} \cdot \sqrt{2} \cdot \frac{T(\frac{n}{2})}{T(\frac{n-1}{2})}$$

$$\Rightarrow \quad C = \sqrt{\frac{n-1}{2}} \cdot \frac{T(\frac{n-1}{2})}{T(\frac{n}{2})}, \quad \text{That's } c \text{ in: } g(S^2) = C \cdot \sqrt{S^2}$$

5. CB 5.15

5.15 Establish the following recursion relations for means and variances. Let $\bar{X}_n$ and S_n^2 be the mean and variance, respectively, of $X_1, \ldots, X_n$. Then suppose another observation, X_{n+1}, becomes available. Show that

(a) $\bar{X}_{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}$.

(b) $nS_{n+1}^2 = (n-1)S_n^2 + \left(\frac{n}{n+1}\right)(X_{n+1} - \bar{X}_n)^2$.

(a)

$$\bar{X}_{n+1} = \frac{\sum_{j=1}^{n+1} X_i}{n+1} = \frac{\sum_{i=1}^{n} X_i}{n+1} + \frac{X_{n+1}}{n+1}$$

$$\bar{X}_n = \frac{\sum_{j=1}^{n} X_i}{n} \Rightarrow \sum_{i=1}^{n} X_i = n\bar{X}_n$$

$$\Rightarrow \bar{X}_{n+1} = \frac{n\bar{X}_n + X_{n+1}}{n+1}$$

(b)

$$n \cdot S_{n+1}^2 = \sum_{i=1}^{n+1} X_i^2 - (n+1)\bar{X}_{n+1}^2 \quad ①$$

$$(n-1) S_n^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2 \quad ②$$

$$① - ② = X_{n+1}^2 - (n+1)\bar{X}_{n+1}^2 + n\bar{X}_n^2$$

$$(n+1)\bar{X}_{n+1}^2 = \frac{n^2\bar{X}_n^2 + 2n\bar{X}_n X_{n+1} + X_{n+1}^2}{n+1}$$

$$-(n+1)\bar{X}_{n+1}^2 + n\bar{X}_n^2 = \frac{n\bar{X}_n^2 - 2n\bar{X}_n X_{n+1} - X_{n+1}^2}{n+1} \quad ③$$

$$③ + X_{n+1}^2 = \frac{n}{n+1}\left(\bar{X}_n^2 - 2\bar{X}_n X_{n+1} + X_{n+1}^2\right) = \frac{n}{n+1}(X_{n+1} - \bar{X}_n)^2$$

④

① = ② + ④.   Q.Z.D.

6. CB 5.21

5.21 What is the probability that the larger of two continuous iid random variables will exceed the population median? Generalize this result to samples of size n.

$P(\max(X_1, X_2) \geq \text{median}) = 1 - P(X_1 \leq M, X_2 \leq M)$

Since it's i.i.d.: $P(X_1 \leq M, X_2 \leq M) = P(X_1 \leq M) \cdot P(X_2 \leq M)$

$= 1/4$ $\Rightarrow P(\max(X_1, X_2) \geq \text{median}) = 1 - 1/4 = 3/4$

For n samples: $P(\max(X_1, \cdots, X_n) \geq M) = 1 - \prod_{i=1}^{n} P(X_i \leq M)$

$= 1 - 1/2^n$

7. CB 5.23

5.23 Let $U_i, i = 1, 2, \ldots,$ be independent uniform$(0,1)$ random variables, and let X have distribution

$$P(X = x) = \frac{c}{x!}, \quad x = 1, 2, 3, \ldots,$$

where $c = 1/(e-1)$. Find the distribution of

$$Z = \min\{U_1, \ldots, U_X\}.$$

(*Hint:* Note that the distribution of $Z|X = x$ is that of the first-order statistic from a sample of size x.)

$P(Z > z | X = x) = P(U_1 > z) P(U_2 > z) \cdots P(U_x > z)$

$= (1-z)^x$

$P(Z \leq z | X = x) = 1 - (1-z)^x \Rightarrow f_{Z|X}(z, x) = \dfrac{\partial P(Z \leq z | X = x)}{\partial z}$

$= \dfrac{\partial(-(1-z)^x)}{\partial z} = x(1-z)^{x-1}$

$$f_{X,Z}(x,z) = x(1-z)^{x-1} \cdot \frac{c}{x!}$$

$$= \frac{(1-z)^{x-1}}{(x-1)!} \cdot c$$

$$f_Z(z) = \sum_{x=1}^{\infty} \frac{(1-z)^{x-1}}{(x-1)!} \cdot c$$

$$= \sum_{t=0}^{\infty} \frac{(1-z)^t}{t!} \cdot c$$

$$= c \cdot e^{1-z} \cdot \sum_{t=0}^{\infty} \cdot \underbrace{\frac{(1-z)^t}{t!} \cdot e^{-(1-z)}}_{\text{Poisson pdf, sums to 1}}$$

$$= c \cdot e^{1-z}$$

$$= \frac{e^{1-z}}{e-1}$$

8. CB 5.31

5.31 Suppose $\bar{X}$ is the mean of 100 observations from a population with mean μ and variance $\sigma^2 = 9$. Find limits between which $\bar{X} - \mu$ will lie with probability at least .90. Use both Chebychev's Inequality and the Central Limit Theorem, and comment on each.

By CLT: $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[D]{} N(0, \sigma^2)$

$\Rightarrow$ $\bar{X}_n$ approximately follows : $N(0, \frac{\sigma^2}{n})$, i.e. $N(0, \frac{9}{100})$

$sd = \frac{3}{10} = 0.3$. use: $\begin{cases} \text{qnorm}: (0.05, mu=0, sd=0.3) = -0.4935 \\ \text{qnorm}: (0.95, mu=0, sd=0.3) = 0.4935 \end{cases}$

so, the range calculate by CLT is: $[-0.4935, \quad 0.4935]$

Chebysheve's Inequality:

$E(\bar{X}_n) = \mu$, $sd(\bar{X}_n) = \frac{\sigma}{\sqrt{n}} = 0.3$

$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$ let $\frac{1}{k^2} = 0.1 \Rightarrow k = \sqrt{10}$

$\Rightarrow$ The range calculated by Chebysheve's is:

$(-\sqrt{10} \cdot 0.3, \sqrt{10} \cdot 0.3)$, $\sqrt{10} \approx 3.16$. $\sqrt{10} \times 0.3 = 0.9487$

Comments:

1. If population follows normal distribution, it would be perfect to use CLT method since it perfectly follows normal distribution.
2. Chebysheve's inequality uses less information. but useful when population is small and not that normal.

9. Let $Z_1, \ldots, Z_n$ be a random sample from a population that is normally distributed with mean μ and variance σ^2. What is the approximate distribution of $(\bar{Z}_n)^4$?

Since Z_i's are i.i.d. $N(\mu, \sigma^2)$, $\bar{Z}_n \sim N(\mu, \frac{\sigma^2}{n})$

i.e. $\sqrt{n}(\bar{Z}_n - \mu) \sim N(0, \sigma^2)$. $g(x) = x^4$

$g'(\mu) = 4\mu^3$, by delta method:

$$\sqrt{n}\left((\bar{Z}_n)^4 - \mu^4\right) \xrightarrow[D]{} N\left(0, \sigma^2 \cdot (4\mu^3)^2\right)$$

10. CB 5.44 part (a) and (b)

5.44 Let $X_i, i = 1, 2, \ldots,$ be independent Bernoulli(p) random variables and let $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) Show that $\sqrt{n}\,(Y_n - p) \to n\,[0, p(1-p)]$ in distribution.

(b) Show that for $p \neq 1/2$, the estimate of variance $Y_n(1-Y_n)$ satisfies $\sqrt{n}[Y_n(1-Y_n) - p(1-p)] \to n\left[0, (1-2p)^2 p(1-p)\right]$ in distribution.

(a) $E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p \quad Var(X_i) = p(1-p)$

Since X_i's are i.i.d., by CLT:

$$\sqrt{n}(\bar{X}_n - p) \xrightarrow[D]{} N(0, p(1-p))$$

$g'(x) \neq 0,$
$x \neq \frac{1}{2}$

(b). let $g(x) = x(1-x)$. $\quad g'(x) = 1-2x \quad$ by delta method:

$$\sqrt{n}\left(Y_n(1-Y_n) - p(1-p)\right) \xrightarrow[D]{} N\left(0, p(1-p)(1-2p)^2\right)$$