# Homework # 8 *Statistics 134, Bandyopadhyay, Spring 2014*

**3.3.24** a)

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(S_2 = x)$ | .0625 | .25 | .375 | .25 | .0625 |

b) $E(S_{50}) = 50$ and $Var(S_{50}) = 50(.5) = 25$. Thus by normal approximation,

$$P(S_{50} = 50) \approx \Phi\left(\frac{50.5 - 50}{5}\right) - \Phi\left(\frac{49.5 - 50}{5}\right) = 0.0797$$

c) $S_n = X_1 + \cdots + X_n$ where the $X_i$ are independent, each with binomial $(2, 1/2)$ distribution. It follows that $S_n$ has binomial $(2n, 1/2)$ distribution:

$$P(S_n = k) = \binom{2n}{k}\left(\frac{1}{2}\right)^{2n}$$

**3.3.30** $D_i^2$ takes the values 0, 1, 4, 9, 16, 25, 36, 49, 64, 81 with equal probability, so the $X_i$ are independent with common distribution

| $x$ | 0 | 1 | 4 | 5 | 6 | 9 |
|---|---|---|---|---|---|---|
| $P(X_i = x)$ | .1 | .2 | .2 | .1 | .2 | .2 |

So $E(X_i) = 9/2$ and $Var(X_i) = 9.05 \Longrightarrow SD(X_i) = 3.008$.

a) By the law of averages, you expect $\bar{X}_n$ to be close to $E(\bar{X}_n) = 4.5$ for large $n$. So predict 4.5.

b) $P(|\bar{X}_n - 4.5| > \epsilon)$
$= P\left(\frac{\sqrt{n}|\bar{X}_n - 4.5|}{3.008} > \frac{\sqrt{n}\epsilon}{3.008}\right) \approx P\left(|Z| > \frac{\sqrt{n}\epsilon}{3.008}\right) = 2P\left(Z > \frac{\sqrt{n}\epsilon}{3.008}\right)$,
where $Z$ has standard normal distribution. For $n = 10000$, we need $\epsilon$ such that
$P\left[Z > \frac{\sqrt{n}\epsilon}{3.008}\right] = \frac{1}{400} = 0.0025 \Longrightarrow \frac{\sqrt{n}\epsilon}{3.008} = 2.81$,
therefore $\epsilon = 2.81 \times 3.008/100 = 0.085$.

c) Need $n$ such that $P(|\bar{X}_n - 4.5| \leq 0.01) \geq 0.99$ i.e.,
$P\left[|Z| \leq \frac{\sqrt{n}\times 0.01}{3.008}\right] \geq 0.99 \Longrightarrow \frac{\sqrt{n}}{300.8} \geq 2.58$
therefore $n \geq 602276$.

d) We have calculated $E(X_i) = 9/2$ and $Var(X_i) = 9.05$. From the previous problem, we have $E(D_i) = 9/2$ and $Var(D_i) = 33/4 = 8.25$. Since $D_i$ has smaller variance than does $X_i$, the value of $\bar{D}_n$ can be predicted more accurately.

e) Since $E(\bar{X}_{100}) = 4.5$, you should predict the first digit of $\bar{X}_{100}$ to be 4. The chance of being correct is

$P(4 \leq \bar{X}_{100} < 5) \approx P\left(\left|\frac{\sqrt{100}(\bar{X}_{100}-4.5)}{3.008}\right| \leq \frac{\sqrt{100}}{2\times 3.008}\right) \approx P(|Z| \leq 1.66) = 0.903.$

**4.1.2** a)

$$\int_1^\infty \frac{c}{x^4}dx = \frac{-c}{3x^3}\Big|_1^\infty = \frac{c}{3}$$

and since $f(x)$ is a density function, it must integrate to 1, so $c = 3$.

b)

$$E(X) = \int_1^\infty x\frac{3}{x^4}dx = \frac{-3}{2x^2}\Big|_1^\infty = \frac{3}{2}$$

c)

$$E(X^2) = \int_1^\infty x^2\frac{3}{x^4}dx = \frac{-3}{x}\Big|_1^\infty = 3$$

Thus $Var(X) = E(X^2) - (E(X))^2 = 3 - \frac{9}{4} = \frac{3}{4}$.

**4.1.4** a) We know that $\int_0^1 cx^2(1-x)^2dx = 1$, and

$$\int_0^1 cx^2(1-x)^2dx = c\int_0^1 (x^2 - 2x^3 + x^4)dx = c\left(\frac{x^3}{3} - \frac{x^4}{2} + \frac{x^5}{5}\right)\Big|_0^1 = \frac{c}{30}$$

so $c = 30$.

b)

$$E(X) = \int_0^1 30x^3(1-x)^2dx = 30\int_0^1 (x^3 - 2x^4 + x^5)dx = 30\left(\frac{1}{4} - \frac{2}{5} + \frac{1}{6}\right) = \frac{1}{2}$$

c)

$$E(X^2) = \int_0^1 30x^4(1-x)^2dx = 30\int_0^1 (x^4 - 2x^5 + x^6)dx = 30\left(\frac{1}{5} - \frac{1}{3} + \frac{1}{7}\right) = \frac{2}{7}$$

and so

$$Var(X) = \frac{2}{7} - \frac{1}{4} = \frac{1}{28}$$

**4.1.12** These are best done by folding the shape over the horizontal axis. You don't need to find an equation for the density if you can give a precise description of the shape of its graph.

a) Possible values of $X$: the interval $[-2, 2]$. The graph of the density will be a triangle whose base is $[-2, 2]$ and whose peak is over $x = 0$. For the area to be 1 the height must be 1/2. Done. If you want to write a formula:

For $-2 \leq x \leq 2$, $f(x)dx = P(X \in dx) = \frac{2\times(2-|x|)dx}{4\times(\frac{1}{2}\times 2\times 2)} = \frac{1}{4}(2-|x|)dx$,

so for $x \in [-2, 2]$, $f(x) = (2-|x|)/4$. Elsewhere $f(x) = 0$ .

b) Possible values of $X$: the interval $[-2, 1]$. Here there isn't even any folding to do. The graph of the density is a triangle whose base is $[-2, 1]$ and whose peak is over $x = 0$. For the area to be 1 the height must be $2/3$. For those who like equations:

If $-2 \leq x < 0$, then
$f(x)dx = P(X \in dx) = \frac{(2+x)dx}{\frac{1}{2}\times 3\times 2} = \frac{1}{3}(2 + x)dx$, so $f(x) = \frac{1}{3}(2 + x)$.

If $0 \leq x \leq 1$, then $f(x)dx = \frac{2(1-x)dx}{\frac{1}{2}\times 3\times 2}$, so $f(x) = \frac{2}{3}(1 - x)$.

Elsewhere $f(x) = 0$.

c) This one requires folding and a little observation. Possible values of $X : [-1, 2]$.

The density will be linear on $[-1, 0]$, constant on $[0, 1]$, linear on $[1, 2]$. That's a trapezoid.

For the area to be 1, $h$ must satisfy $2h = 1$, or $h = 1/2$.