FUNCTIONS OF RANDOM VARIABLES Solutions

Definition. Let $g(x)$ be a continuous function and let X be a continuous random variable with PDF $f(x)$. Then $E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\,dx$

1. Let X be a continuous random variable with mean 50 and standard deviation 4. Calculate $E\left(\frac{X-50}{4}\right)$ and $\text{Var}\left(\frac{X-50}{4}\right)$. Hint: the given information includes $50 = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$ and $16 = \text{Var}(X) = E(X^2) - 50^2$.

$$E\left(\frac{X-50}{4}\right) = \int_{-\infty}^{\infty}\left(\frac{x-50}{4}\right)f(x)\,dx = \frac{1}{4}\underbrace{\int_{-\infty}^{\infty} x f(x)\,dx}_{E(X)} - \frac{50}{4}\underbrace{\int_{-\infty}^{\infty} f(x)\,dx}_{1} = \frac{1}{4}E(X) - \frac{50}{4} = 0$$

$16 = Var(X) = E(X^2) - 50^2$ so $E(X^2) = 2516$.

$$Var\left(\frac{X-50}{4}\right) = E\left[\left(\frac{X-50}{4}\right)^2\right] - \left[E\left(\frac{X-50}{4}\right)\right]^2 = E\left[\left(\frac{X-50}{4}\right)^2\right]$$

$$= \int_{-\infty}^{\infty}\left(\frac{x-50}{4}\right)^2 f(x)\,dx =$$

$$= \frac{1}{16}\int_{-\infty}^{\infty}(x^2 - 100x + 2500)f(x)\,dx$$

$$= \frac{1}{16}\left[\underbrace{\int_{-\infty}^{\infty} x^2 f(x)\,dx}_{E(X^2)} - 100\underbrace{\int_{-\infty}^{\infty} x f(x)\,dx}_{E(X)} + 2500\underbrace{\int_{-\infty}^{\infty} f(x)\,dx}_{1}\right]$$

$$= \frac{1}{16}\left[2516 - 100(50) + 2500\right] = 1$$

Challenge. Fill in the blanks in the following theorem with expressions involving a, b, μ, and σ.

Theorem 1. *Let X be a random variable with mean μ and variance σ^2 and let a and b be constants.*

i) $E(aX+b) = \boxed{a\mu + b} = aE(X) + b$

ii) $Var(aX+b) = \boxed{a^2\sigma^2} = a^2 Var(X)$

$$E(aX+b) = \int_{-\infty}^{\infty}(ax+b)f(x)\,dx = a\int_{-\infty}^{\infty} x f(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx = aE(X) + b$$

$$Var(aX+b) = E\left[\left(aX+b-(a\mu+b)\right)^2\right] = E\left[a^2(X-\mu)^2\right] = a^2 E\left[(X-\mu)^2\right] = a^2 Var(X)$$

Discrete case similar.

Date: March 3, 2023.

Theorem 2. *Let $X_1, X_2, \ldots, X_n$ be any random variables. Then $E(X_1 + X_1 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$.*

Theorem 3. *Let $X_1, X_2, \ldots, X_n$ be independent random variables. Then $Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$.*

Definition. We say that the random variables $X_1, X_2, \ldots, X_n$ are a **random sample** if they are independent and identically distributed. The **sample size** is n. Their common distribution is called the **population distribution**. Some **sample statistics**:

1. The **sample total**: $T = \sum_{i=1}^{n} X_i = X_1 + X_1 + \cdots + X_n$ $E(T) = n\mu$ $Var(T) = n\sigma^2$
2. The **sample mean**: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_1 + \cdots + X_n)$ $E(\overline{X}) = \mu$ $Var(\overline{X}) = \frac{\sigma^2}{n}$
3. The **sample variance**: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Our goal is to understand the distributions of these sample statistics.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. Calculate the expected value and variance of the sample mean (that's $E(\overline{X})$ and $\mathrm{Var}(\overline{X})$).

$$
\begin{aligned}
E(\overline{X}) &= E\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\
&= \frac{1}{n} E(X_1 + X_2 + \cdots X_n) \\
&= \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right] && \text{by theorem 2} \\
&= \frac{1}{n}\left[\underbrace{\mu + \mu + \cdots + \mu}_{n \text{ times}}\right] && \text{since } E(X_i) = \mu \text{ for } i = 1, 2, \ldots, n \\
&= \frac{1}{n}[n\mu] \\
&= \mu
\end{aligned}
$$

$$
\begin{aligned}
Var(\overline{X}) &= Var\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\
&= \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) \\
&= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + \cdots + Var(X_n)\right] && \text{by theorem 3} \\
&= \frac{1}{n^2}\left[\underbrace{\sigma^2 + \sigma^2 + \cdots + \sigma^2}_{n \text{ times}}\right] && \text{since } Var(X_i) = \sigma^2 \text{ for } i = 1, 2, \ldots, n \\
&= \frac{1}{n^2}[n\sigma^2] \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

Theorem 4 (Central Limit Theorem). *If $X_1, X_2, \ldots, X_n$ comprise a random sample from a population with mean μ and variance σ^2, then the limiting distribution (as $n \to \infty$) of $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is a standard normal distribution.* Note: $\Phi(x) =$ pnorm(x) are both the CDF of the std normal dist.

The CLT tells us that for large sample sizes $\overline{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$. By noticing that $T = n\overline{X}$, we can also say that T approximately normal with mean $n\mu$ and variance $n\sigma^2$ (for large samples). Use these normal approximations to answer the following problems.

3. A certain kind of bicycle spoke has a mean mass of $\mu = 4.2$ g with a standard deviation of $\sigma = 0.4$ g. I intend to build a wheel with $n = 32$ spokes.

a) What is the probability that the total mass of the spokes in my wheel is under 136 g?
b) What is the probability that the total mass of the spokes in my wheel exceeds 130 g?
c) Suppose I build two wheels with these spokes. What is the probability that the total mass of the 64 spokes will be under 272 g?

Note: $T = n\overline{X}$ so $E(T) = nE(\overline{X}) = n\mu$ and $Var(T) = n^2 Var(\overline{X}) = n\sigma^2$.

In this problem $\mu = 4.2$, $\sigma^2 = (0.4)^2$ and $n = 32$. Hence $E(T) = 134.4\,g$ and $Var(T) = 5.12\,g^2$

$T \sim N(134.4, 5.12)$, approximately.

a) $P(T \le 136) \approx \Phi\left(\frac{136-134.4}{\sqrt{5.12}}\right) = \Phi(0.7071) \approx 0.7602$

b) $P(T \ge 130) \approx 1-\Phi\left(\frac{130-134.4}{\sqrt{5.12}}\right) = 1-\Phi(-1.9445) \approx 0.9741$

c) Now $n = 64$, so the variance is now $64(0.4)^2 = 10.24$ and the mean is 268.8

$P(T \le 272) \approx \Phi\left(\frac{272-268.8}{\sqrt{10.24}}\right) = \Phi(1) \approx 0.8413$

4. Suppose you roll a fair 6-sided die 50 times. Estimate the probability that the mean of all the rolls is at least 4. Hint: think of this as a random sample of size 50 from an underlying population with the mean and variance of a single roll of the die (you may need to calculate this mean and variance).

The mean of one roll is $\frac{7}{2}$

The variance of one roll is: $\frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12}$

$\frac{1+4+9+16+25+36}{6} = \frac{91}{6}$

Thus $\overline{X} \sim N\left(\frac{7}{2}, \frac{7}{120}\right)$ and $P(\overline{X} \ge 4) \approx 1-\Phi\left(\frac{4-7/2}{\sqrt{7/120}}\right) \approx 1-\Phi(2.0702)$

≈ 0.0192

5. If the underlying population is normally distributed, then $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is a standard normal random variable (not just approximately standard normal), so our calculations are the same, but our answers are exact and **all sample sizes work**. Suppose we have a normally distributed population with mean 100 and standard deviation 10. Let $\overline{X}$ be the mean of a random sample of size n. Calculate $P(98 < \overline{X} < 102)$ for each sample size.

a) $n = 1$
b) $n = 25$
c) $n = 100$
d) $n = 10000$

$$P(98 < \overline{X} < 102) = P\left(\frac{98-100}{10/\sqrt{n}} < Z < \frac{102-100}{10/\sqrt{n}}\right) = \Phi(0.2\sqrt{n}) - \Phi(-0.2\sqrt{n})$$

n	$P(98 < \overline{X} < 102)$
1	0.1585194
25	0.6826895
100	0.9544997
1000	1

6. Binomial distributions show up when we count the number of successes in a random sample. For example, suppose you roll a fair 6-sided die 64 times. Let T be the number of sixes you roll.

a) What is the (exact) distribution of T?
b) Calculate $P(T \geq 12)$.
c) The CLT says that T is approximately normal with mean 64(1/6) and variance 64(1/6)(5/6). Use this normal approximation to estimate $P(T \geq 12)$. How does this answer compare with the exact answer you found in part b?

$T \sim \text{binom}(64, 1/6)$ and T is approx $N\left(\frac{32}{3}, \frac{4}{3}\sqrt{5}\right)$

$P(T \geq 12) = 1 - \text{pbinom}(11, 64, 1/6) \approx 0.3768897$

$P(T \geq 12) \approx 1 - \text{pnorm}(12, 64/6, \text{sqrt}(5)*4/3) \approx 0.3273604$

With correction for continuity: $P(T \geq 12) \approx 1 - \text{pnorm}(11.5, 64/6, \text{sqrt}(5)*4/3)$

≈ 0.3899273 ← better

OR

$1 - \text{pnorm}\left(\frac{12 - 64/6}{\sqrt{5}(4/3)}\right)$

0.4447163