Chapter 3

Discrete Random Variables & Distributions

Detailed Solutions

3.1 Basic Concept

3.1 Validating PMF

Problem: Determine the constant c for valid PMFs.
Solution:

For a function to be a valid Probability Mass Function (PMF), the sum of probabilities over the entire support must equal 1: $\sum p(x) = 1$.

(a) $p(x) = c(x^2 + 1)$ for $x = 0, 1, 2, 3$.

$$\begin{aligned}
\sum_{x=0}^{3} c(x^2+1) &= 1 \\
c[(0^2+1)+(1^2+1)+(2^2+1)+(3^2+1)] &= 1 \\
c[1+2+5+10] &= 1 \\
18c = 1 &\implies c = \frac{1}{18}
\end{aligned}$$

(b) $p(x) = c \cdot (1/2)^x$ for $x = 1, 2, 3, \ldots$ (Geometric Series).

$$\begin{aligned}
\sum_{x=1}^{\infty} c\left(\frac{1}{2}\right)^x &= 1 \\
c\left[\frac{1/2}{1-1/2}\right] &= 1 \quad (\text{Using } S = \frac{a}{1-r}) \\
c\left[\frac{0.5}{0.5}\right] &= 1 \implies c = 1
\end{aligned}$$

(c) $p(x) = c\binom{5}{x}$ for $x = 0, \ldots, 5$.

$$c\sum_{x=0}^{5}\binom{5}{x} = 1$$
$$c \cdot 2^5 = 1 \quad \text{(Sum of binomial coeffs} = 2^n\text{)}$$
$$32c = 1 \implies c = \frac{1}{32}$$

(a) $c = 1/18$ (b) $c = 1$ (c) $c = 1/32$

3.2 Distribution Identification

Problem: Identify the distribution.
Solution:

(a) **Binomial:** Fixed number of independent trials ($n = 10$), two outcomes (Head/-Tail).

(b) **Poisson:** Counting events (cars) in a fixed interval of time (1 hour).

(c) **Geometric:** Counting number of trials *until* the first success occurs.

(d) **Hypergeometric:** Sampling without replacement from a finite population with two groups (defective/good).

(e) **Negative Binomial:** Counting trials until a specific number of successes ($r = 5$ errors) occurs.

(a) Binomial (b) Poisson (c) Geometric (d) Hypergeometric (e) Negative Binomial

3.2 Intermediate

3.3 Visualizing the CDF

Problem: Analyze step function graph of $F(x)$.
Solution:

(a) **Find PMF:** The probability at x, $P(X = x)$, is the magnitude of the "jump" in the CDF at x.

$$\begin{aligned}
P(X = 0) &= F(0) - F(0^-) = 0.2 - 0 = 0.2 \\
P(X = 1) &= F(1) - F(1^-) = 0.5 - 0.2 = 0.3 \\
P(X = 2) &= F(2) - F(2^-) = 0.9 - 0.5 = 0.4 \\
P(X = 3) &= F(3) - F(3^-) = 1.0 - 0.9 = 0.1
\end{aligned}$$

(b) **Calculate Probability:**

$$\begin{aligned}
P(0.5 < X \leq 2.5) &= F(2.5) - F(0.5) \\
&= 0.9 - 0.2 \quad \text{(read from graph)} \\
&= 0.7
\end{aligned}$$

Alternatively, sum PMF for $x \in \{1, 2\}$: $0.3 + 0.4 = 0.7$.

(c) **Expected Value:**

$$\begin{aligned}
E[X] &= \sum x \cdot p(x) \\
&= (0)(0.2) + (1)(0.3) + (2)(0.4) + (3)(0.1) \\
&= 0 + 0.3 + 0.8 + 0.3 \\
&= 1.4
\end{aligned}$$

(a) $P(0) = 0.2, P(1) = 0.3, P(2) = 0.4, P(3) = 0.1$
(b) 0.7
(c) $E[X] = 1.4$

3.4 Binomial: Quality Control

Problem: $n = 20$, $p = 0.05$ (defective).
Solution:

Let $X \sim B(20, 0.05)$. PMF: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

(a) **Exactly 2 defectives:**

$$\begin{aligned}
P(X = 2) &= \binom{20}{2} (0.05)^2 (0.95)^{18} \\
&= 190 \times 0.0025 \times 0.3972 \\
&\approx 0.1887
\end{aligned}$$

(b) **At least 1 defective:**

$$\begin{aligned}
P(X \geq 1) &= 1 - P(X = 0) \\
&= 1 - \binom{20}{0}(0.05)^0(0.95)^{20} \\
&= 1 - (1 \times 1 \times 0.3585) \\
&= 0.6415
\end{aligned}$$

(c) **Expectation and Variance:**

$$\begin{aligned}
E[X] &= np = 20(0.05) = 1.0 \\
Var(X) &= np(1-p) = 20(0.05)(0.95) = 0.95
\end{aligned}$$

(a) 0.1887
(b) 0.6415
(c) Mean = 1, Variance = 0.95

3.5 Hypergeometric vs. Binomial

Problem: Pop $N = 50$, Defective $K = 10$, Sample $n = 5$. Find $P(X = 1)$.
Solution:

(a) **Hypergeometric (Exact):**

$$\begin{aligned}
P(X = 1) &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\binom{10}{1}\binom{40}{4}}{\binom{50}{5}} \\
&= \frac{10 \times 91,390}{2,118,760} = \frac{913,900}{2,118,760} \approx 0.4313
\end{aligned}$$

(b) **Binomial Approximation:** Set $p = K/N = 10/50 = 0.2$.

$$\begin{aligned}
P(X = 1) &\approx \binom{5}{1}(0.2)^1(0.8)^4 \\
&= 5 \times 0.2 \times 0.4096 = 0.4096
\end{aligned}$$

Reasoning: The approximation (0.4096) is reasonably close to the exact value (0.4313) but strictly speaking, the rule of thumb $n/N < 0.1$ is barely met ($5/50 = 0.1$). Binomial assumes replacement (constant probability), while Hypergeometric accounts for the changing probability as items are removed.

(a) Hypergeometric: ≈ 0.4313
(b) Binomial: ≈ 0.4096. Close approximation as $n/N \leq 0.1$.

3.6 Poisson: Web Server Traffic

Problem: Rate $\lambda = 5$ req/sec.
Solution:

(a) **Exactly 3 requests in 1 second:** Here $\lambda_t = 5$.

$$P(X = 3) = \frac{e^{-5}5^3}{3!} = \frac{0.006738 \times 125}{6} \approx 0.1404$$

(b) **No requests in 2 seconds:** New rate $\Lambda = \lambda \times t = 5 \times 2 = 10$.

$$P(X = 0) = \frac{e^{-10}10^0}{0!} = e^{-10} \approx 0.000045$$

(c) **Mode:** For Poisson, the mode is at $\lfloor \lambda \rfloor$. Since $\lambda = 5$ is an integer, the modes are 4 and 5. Usually, we refer to the peak at λ.

(a) 0.1404
(b) 4.54×10^{-5}
(c) 5 (and 4)

3.7 Negative Binomial: Oil Drilling

Problem: $p = 0.2$ (Success = Oil). Independent trials.
Solution:

(a) **First discovery on 4th drill:** This is Geometric distribution ($r = 1$) or NegBinom with $k = 4, r = 1$.

$$\begin{aligned} P(X = 4) &= (1 - p)^{4-1}p \\ &= (0.8)^3(0.2) = 0.512 \times 0.2 = 0.1024 \end{aligned}$$

(b) **Third discovery on 10th drill:** NegBinom with $r = 3$ successes in $x = 10$ trials. Formula: $P(X = x) = \binom{x-1}{r-1}p^r(1-p)^{x-r}$.

$$\begin{aligned} P(X = 10) &= \binom{10-1}{3-1}(0.2)^3(0.8)^{10-3} \\ &= \binom{9}{2}(0.008)(0.8)^7 \\ &= 36 \times 0.008 \times 0.2097 \\ &\approx 0.0604 \end{aligned}$$

(c) **Expected number of drills for 3 successes:**

$$E[X] = \frac{r}{p} = \frac{3}{0.2} = 15$$

(a) 0.1024
(b) 0.0604
(c) 15 drills

3.8 Expectation and Variance Calculation

Problem: $x \in \{-2, 0, 1, 3\}$ with probs $\{0.1, 0.2, 0.4, 0.3\}$.
Solution:

(a) **E[X] and Var(X):**

$$\begin{aligned} E[X] &= (-2)(0.1) + (0)(0.2) + (1)(0.4) + (3)(0.3) \\ &= -0.2 + 0 + 0.4 + 0.9 = 1.1 \\ E[X^2] &= (-2)^2(0.1) + 0 + (1)^2(0.4) + (3)^2(0.3) \\ &= 4(0.1) + 0.4 + 9(0.3) = 0.4 + 0.4 + 2.7 = 3.5 \\ Var(X) &= E[X^2] - (E[X])^2 = 3.5 - (1.1)^2 = 3.5 - 1.21 = 2.29 \end{aligned}$$

(b) **Linear Transformations:**

$$\begin{gathered} E[2X+5] = 2E[X] + 5 = 2(1.1) + 5 = 7.2 \\ E[X^2] = 3.5 \quad \text{(Calculated above)} \end{gathered}$$

(c) **Variance Property:**

$$Var(2X+5) = 2^2 Var(X) = 4(2.29) = 9.16$$

(a) $E[X] = 1.1, Var(X) = 2.29$
(b) $E[2X+5] = 7.2, E[X^2] = 3.5$
(c) 9.16

3.9 Custom Discrete RV (Infinite Series)

Problem: $P(X = k) = c \cdot k \cdot (1/3)^k$.
Solution:

Let $r = 1/3$. The PMF is $c \cdot k \cdot r^k$.

(a) **Find c:** Sum must be 1.

$$\sum_{k=1}^{\infty} ckr^k = c\left[\frac{r}{(1-r)^2}\right]$$
$$c\left[\frac{1/3}{(2/3)^2}\right] = c\left[\frac{1/3}{4/9}\right] = c\left[\frac{1}{3} \cdot \frac{9}{4}\right] = c\frac{3}{4}$$
$$c\frac{3}{4} = 1 \implies c = \frac{4}{3}$$

(b) **Find F(2):**

$$F(2) = P(X = 1) + P(X = 2)$$
$$= \frac{4}{3}\left[1\left(\frac{1}{3}\right)^1 + 2\left(\frac{1}{3}\right)^2\right]$$
$$= \frac{4}{3}\left[\frac{1}{3} + \frac{2}{9}\right] = \frac{4}{3}\left[\frac{3}{9} + \frac{2}{9}\right] = \frac{4}{3} \cdot \frac{5}{9} = \frac{20}{27}$$

(c) **Find E[X]:**

$$E[X] = \sum_{k=1}^{\infty} k \cdot P(k) = \sum_{k=1}^{\infty} k \cdot (ckr^k) = c\sum k^2 r^k$$
$$= c\left[\frac{r(1+r)}{(1-r)^3}\right] = \frac{4}{3}\left[\frac{(1/3)(4/3)}{(2/3)^3}\right]$$
$$= \frac{4}{3}\left[\frac{4/9}{8/27}\right] = \frac{4}{3}\left[\frac{4}{9} \cdot \frac{27}{8}\right] = \frac{4}{3} \cdot \frac{3}{2} = 2$$

(a) $c = 4/3$
(b) $F(2) = 20/27 \approx 0.74$
(c) $E[X] = 2$

3.10 Binomial Approximation by Poisson

Problem: $n = 10{,}000$, $p = 0.0002$. Find $P(X = 3)$.
Solution:

(a) **Exact Binomial:**

$$P(X = 3) = \binom{10000}{3}(0.0002)^3(0.9998)^{9997}$$

(b) **Poisson Approximation:** Calculate $\lambda = np = 10000 \times 0.0002 = 2$.

$$\begin{aligned} P(X = 3) &\approx \frac{e^{-2}2^3}{3!} \\ &= \frac{0.1353 \times 8}{6} \approx 0.1804 \end{aligned}$$

(c) **Mean and Variance Comparison:**

$$\begin{aligned} \text{Binomial Mean} &= np = 2 \\ \text{Poisson Mean} &= \lambda = 2 \quad \text{(Identical)} \\ \text{Binomial Var} &= np(1-p) = 2(0.9998) = 1.9996 \\ \text{Poisson Var} &= \lambda = 2 \end{aligned}$$

Difference is 0.0004. Since $p \to 0$, $1 - p \to 1$, making $np(1-p) \approx np$.

(a) $\binom{10000}{3}(0.0002)^3(0.9998)^{9997}$
(b) 0.1804
(c) Difference is negligible (0.0004), justifying approximation.

3.3 Challenge

3.11 Geometric Mean Derivation

Problem: Prove $E[X] = 1/p$ for Geometric distribution.
Solution:

Definition of expectation: $E[X] = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1}p$. Let $q = 1-p$. Since $0 < q < 1$:

$$\begin{aligned} E[X] &= p\sum_{k=1}^{\infty} kq^{k-1} \\ &= p\sum_{k=1}^{\infty} \frac{d}{dq}(q^k) \\ &= p\frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) \quad \text{(Geometric Series } \sum q^k = \frac{q}{1-q}\text{)} \\ &= p\frac{d}{dq}\left(\frac{q}{1-q}\right) \end{aligned}$$

Using Quotient Rule $\left(\frac{u}{v}\right)' = \frac{u'v-uv'}{v^2}$:

$$\frac{d}{dq}\left(\frac{q}{1-q}\right) = \frac{1(1-q)-q(-1)}{(1-q)^2} = \frac{1-q+q}{(1-q)^2} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$$

Substitute back:

$$E[X] = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

$\boxed{\text{Proven.}}$

3.12 Poisson as a Limit of Binomial

Problem: Show limit of Binomial is Poisson as $n \to \infty, p \to 0, np = \lambda$.
Solution:

$$\begin{aligned} P(X=k) &= \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\ \text{Sub } p = \lambda/n : &= \frac{n(n-1)\dots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\to 1} \cdot \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \cdot \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1} \end{aligned}$$

As $n \to \infty$:

1. The fraction $\frac{n(n-1)\dots}{n^k} = 1(1-\frac{1}{n})\dots \to 1$.

2. The definition of exponential: $\lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$.
3. The term $(1-\frac{\lambda}{n})^{-k} \to (1-0)^{-k} = 1$.

Thus, expression becomes $\frac{\lambda^k e^{-\lambda}}{k!}$.

Proven.

3.13 Lack of Memory (Geometric)

Problem: Prove $P(X > s+t|X > s) = P(X > t)$.
Solution:

For Geometric distribution, $P(X > k) = (1-p)^k = q^k$.

$$\begin{aligned} LHS &= \frac{P(X > s+t \cap X > s)}{P(X > s)} \\ &= \frac{P(X > s+t)}{P(X > s)} \quad \text{(Since } X > s+t \implies X > s) \\ &= \frac{q^{s+t}}{q^s} \\ &= q^t \\ &= P(X > t) = RHS \end{aligned}$$

Physical Meaning: If you have flipped a coin s times and got tails, the probability of needing t MORE flips to get a head is the same as if you just started. The coin has no memory of past failures.

3.14 Unknown Distribution Bounds (Chebyshev)

Problem: $\mu = 5.0, \sigma = 0.1$.
Solution:

Chebyshev's Inequality: $P(|X-\mu| < k\sigma) \ge 1 - \frac{1}{k^2}$.

1. **Range 4.8 to 5.2:** Deviation is $|5.2 - 5.0| = 0.2$. $k\sigma = 0.2 \implies k(0.1) = 0.2 \implies k = 2$.

$$P(4.8 < X < 5.2) \ge 1 - \frac{1}{2^2} = 1 - 0.25 = 0.75$$

2. **At least 75%:** This corresponds to $1 - 1/k^2 = 0.75$, so $k = 2$. Range is $\mu \pm 2\sigma = 5.0 \pm 0.2 = [4.8, 5.2]$.

(1) Probability ≥ 0.75 (2) Range $[4.8, 5.2]$

3.15 Chebyshev vs. Normal

Problem: $\mu = 100, \sigma = 10$. Range $(80, 120)$.
Solution:

Range is $\mu \pm 20$, so $k\sigma = 20 \implies k = 2$.

1. **Chebyshev:** $P \geq 1 - 1/2^2 = 0.75$.
2. **Normal:** $P(\mu - 2\sigma < X < \mu + 2\sigma)$. From Z-table, approx 95.44% (0.9544).
3. **Comparison:** Chebyshev (75%) is much lower than Normal (95%). Chebyshev is a "loose" bound because it must hold true for *any* distribution (even very weird ones), whereas Normal makes a strong assumption about the shape.

$\boxed{\text{Chebyshev: } \geq 0.75\text{, Normal: } \approx 0.954.}$

3.16 Derivation from Markov's Inequality

Problem: Prove Chebyshev using Markov.
Solution:

Markov's Inequality: $P(Y \geq a) \leq \frac{E[Y]}{a}$. We want to bound $P(|X - \mu| \geq k\sigma)$.

1. Define $Y = (X - \mu)^2$. Since it is squared, $Y \geq 0$.
2. The event $|X - \mu| \geq k\sigma$ is equivalent to $(X - \mu)^2 \geq k^2\sigma^2$.
3. Apply Markov with $a = k^2\sigma^2$:

$$P((X - \mu)^2 \geq k^2\sigma^2) \leq \frac{E[(X - \mu)^2]}{k^2\sigma^2}$$

4. Recall that $E[(X - \mu)^2] = Var(X) = \sigma^2$.

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

$\boxed{\text{Proven.}}$

3.4 Application

3.17 Binomial Convergence Simulation

Problem: Python code for visualization.
Solution:

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom, norm

def plot_binom_convergence():
    p = 0.5
    n_values = [10, 30, 100, 1000]

    plt.figure(figsize=(12, 8))

    for i, n in enumerate(n_values):
        plt.subplot(2, 2, i+1)

        # Binomial Data
        x = np.arange(0, n+1)
        pmf = binom.pmf(x, n, p)
        plt.bar(x, pmf, alpha=0.5, label=f'Binom(n={n})')

        # Normal Approximation
        mu = n * p
        sigma = np.sqrt(n * p * (1 - p))
        x_norm = np.linspace(0, n, 1000)
        pdf = norm.pdf(x_norm, mu, sigma)
        plt.plot(x_norm, pdf, 'r-', linewidth=2, label='Normal Approx')

        plt.title(f'n = {n}')
        plt.legend()

    plt.tight_layout()
    plt.show()

# Run the function
# plot_binom_convergence()
```

3.18 Poisson vs. Real Data

Problem: Data = [0, 1, 0, 2, 1, 0, 5, 1, 0, 2, 1, 0, 0, 3, 1].
Solution:

(a) **Calculate Mean:** Sum $= 0+1+0+2+1+0+5+1+0+2+1+0+0+3+1 = 17$. Count $n = 15$. $\bar{x} = 17/15 \approx 1.133$. Let $\lambda = 1.133$.

(b) **Theoretical Probabilities ($k = 0, 1, 2, 3$):** Using $P(k) = \frac{e^{-1.133}(1.133)^k}{k!}$:

- $k = 0 : e^{-1.133} \approx 0.322$
- $k = 1 : 0.322 \times 1.133 \approx 0.365$

- $k = 2 : 0.365 \times 1.133/2 \approx 0.207$
- $k = 3 : 0.207 \times 1.133/3 \approx 0.078$

(c) **Comparison:** Observed Frequencies (Total 15):

- 0: 6 times $\rightarrow 6/15 = 0.40$
- 1: 5 times $\rightarrow 5/15 = 0.33$
- 2: 2 times $\rightarrow 2/15 = 0.13$
- 3: 1 time $\rightarrow 1/15 = 0.06$

The theoretical values (0.32, 0.36) are reasonably close to observed (0.40, 0.33) given the small sample size.

$\boxed{\lambda \approx 1.13.\ \text{Data follows Poisson trend roughly.}}$