Random Variables

Example: genders of 3 children

<table>
<thead>
<tr>
<th>Gender</th>
<th>BBB</th>
<th>GBB</th>
<th>GGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = No. of Boys</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Y = Difference in no. of each gender</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Y(BBB) = 3
Y(GBG) = 1
Y(GGG) = 0

Probability Mass Function

Space of X: {0, 1, 2, 3}

pmf of X:

- $p(0) = Pr(X = 0) = Pr(\{GGG\}) = 0.125$
- $p(1) = Pr(X = 1) = Pr(\{BBG, GGB, GBB\}) = 0.375$
- $p(2) = Pr(X = 2) = Pr(\{BBB\}) = 0.125$
- $p(3) = Pr(X = 3) = Pr(\{BBB\}) = 0.125$

Properties of pmf:

1. $p(x) \geq 0$
2. $\sum_{x \in X} p(x) = 1$
3. $Pr(X \notin A) = \sum_{x \notin A} p(x)$

Mathematical Expectation

Example: Sic Bo

Toss 3 fair dice

Bet option 1: win $24 if five numbers show up as six; otherwise lose $1
Bet option 2: win $24 if same number on three dice; otherwise lose $1

Expected value of random variable $X$

$E(X) = \sum_{x \in X} xp(x)$

Expected value of $X^2$

$E(X^2) = \sum_{x \in X} x^2 p(x)$

Expected value of $\log(X)$

$E(\log X) = \sum_{x \in X} (\log x) p(x)$

Mathematical Expectation

Example: investment

<table>
<thead>
<tr>
<th>Return on Stock</th>
<th>10%</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>

$E(S) = 9\%$

$Var(S) = 89\%$

$E(B) = 8\%$

$Var(B) = 1.6\%$

Mean and Variance

Example: investment

Population Variance $\sigma^2 = Var(X) = E[(X - \mu)^2]$

$\sigma^2 = Var(X) = E[X^2 - 2\mu X + \mu^2]$

Population Variance $\sigma^2 = Var(X) = E[X^2] - \mu^2$

Mean and Variance

Sample Use relative frequency

Population Use pmf $p(x)$

$\mu = \sum x p(x)$

$\mu = \sum x f(x)$

$\sigma^2 = \sum (x - \mu)^2 f(x)$

$\sigma^2 = \sum (x - \mu)^2 p(x)$

$E(aX + c) = aE(X) + c$

$Var(aX) = a^2 Var(X)$
Mean and Variance
Example: water flow

\[ \text{Pr}(\text{open}) = \text{Pr}(\text{ii open}) = \text{Pr}(\text{iii open}) = 0.8 \]

**Y** = no. of open paths from A to B

\[ Y \in \{0, 1, 2\} \]

\[ \rho(2) = 0.512, \rho(1) = 0.0721 - \rho(1) = 0.4160 - \rho(2) \]

\[ E(Y) = 1.441072 \times [\rho(Y)]^2 = 2.464 \times \text{Var}(Y) = 0.3904 \times (0.44)^2 \times 512 \]

\[ X = 80Y \]

\[ E(X) = 80E(Y) = 115.2 \]

\[ \text{Var}(X) = 80^2 \text{Var}(Y) = 2498.56 \]

Binomial Distribution

**B** = Bernoulli Experiment/Trial

\[ \begin{array}{c|c|c}
\text{Possible outcome} & \text{Probability} \\
\hline
\text{Success} & \pi \\
\text{Failure} & 1 - \pi \\
\hline
\end{array} \]

**X** = No. of successes in **n** independent Bernoulli trials

\[ \text{Binomial Distribution} \]

\[ nCx \times \pi^x \times (1 - \pi)^{n-x} \]

**Poisson Distribution**

**Example:** no. of customers arrived before time **t**

\[ N \sim \text{Poisson} \]

\[ \text{Assumptions:} \]

1. Independence of non-overlapping time intervals
2. Customers won't cluster in a short interval
3. Probability of arrival in a short interval = length of interval

\[ \text{Pr}(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \]
Poisson Distribution

\[ Y \sim \text{P}(\theta) \]

\[ p(y) = \Pr(Y = y) = \frac{e^{-\theta} \theta^y}{y!}, \quad y = 0, 1, 2, \ldots \]

\[ E(Y) = \text{Var}(Y) = \theta \]

Poisson process:

\[ N(t) = \text{no. of occurrences within } [0, t] \]

\[ \lambda = \text{average number of occurrences per unit time} \]

\[ N(t) \sim \text{P}(\lambda t) \]

Example: average 3 phone calls per hour

time unit = 1 hour, \( \lambda = 3 \)

\[ \Pr(2 \text{ phone calls in one hour}) = \Pr(N(1) = 2) = \frac{e^{-3}3^2}{2!} = 0.224 \]

\[ \Pr(\text{at least 2 phone calls in one hour}) = \Pr(N(1) \geq 2) \]

\[ = 1 - \Pr(N(1) < 2) = 1 - (1 - e^{-3}) - \frac{e^{-3}3^0}{0!} = 0.801 \]

\[ \Pr(\text{less than 8 phone calls in 2 hours}) = ? \]

\[ N(2) \sim \text{P}(6) \]

\[ \Pr(N(2) \leq 7) = \sum_{y=0}^{7} \frac{e^{-6}6^y}{y!} = 0.744 \]

Poisson Approximation to Binomial

If \( n \) is large and \( \pi \) is small, binomial can be approximated by Poisson.

\[ b(n, \pi) \equiv \text{P}(n \pi) \]

\[ p(x) = C, \pi^x (1 - \pi)^{n-x} = \frac{e^{-n\pi}(n\pi)^x}{x!} \]

Example: \( X \sim b(8000, 0.001) \)

\[ \Pr(X \leq 7) = \sum_{x=0}^{7} \frac{e^{-8}8^x}{x!} = 0.4530 \]

Continuous Distribution

Recall the relative frequency histogram

Monthly salary of 6000 workers under 25

Density = relative frequency

class width

Monthly salary of 6000 workers under 25

Probability Density Function

Density = \( \lim_{x \to \infty} \frac{\Pr(X \leq x + t) - \Pr(X \leq x)}{t} \)

Monthly salary of all workers under 25

Probability Density Function

Density = \( \lim_{x \to \infty} \frac{\Pr(X \leq x + t) - \Pr(X \leq x)}{t} \)
Distribution Function

\[ f(x) = \frac{1}{a} e^{-x/a}, \quad 0 < x < 2 \]

\[ P(X \leq 0.5) = 0.15625 \]

\[ P(0 \leq X \leq 1.5) = 0.34375 \]

Distribution function \( F(x) = \frac{1}{2} \left( 1 - e^{-x} \right) \)

\[ P(X \leq 0.5) = F(0.5) \]

\[ P(0 \leq X \leq 1.5) = F(1.5) - F(0) \]

Mean and Variance

<table>
<thead>
<tr>
<th></th>
<th>Discrete pmf ( p(x) )</th>
<th>Continuous pdf ( f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \mu )</td>
<td>( \mu = \sum xp(x) )</td>
<td>( \mu = \int x f(x) )</td>
</tr>
<tr>
<td>Variance ( \sigma^2 )</td>
<td>( \sigma^2 = \sum (x - \mu)^2 p(x) )</td>
<td>( \sigma^2 = \int (x - \mu)^2 f(x) )</td>
</tr>
</tbody>
</table>

\[ \text{Var}(X) = E[X^2] - \mu^2 = \int x^2 f(x) dx - \mu^2 \]

Exponential Distribution

Example: \( N(t) \) = no. of customers arrived before time \( t \)

- Time 0
- Assumptions:
  1. Independence of non-overlapping time intervals
  2. Customers won’t cluster in a short interval
  3. Probability of arrival in a short interval = length of interval

\[ Pr(N(t) = y) = \frac{e^{-\lambda t} x^y}{x!}, \quad y = 0,1,2,... \]

Exponential Distribution

Example: \( T = \) waiting time until the first customer arrive

- Time 0
- Assumptions:
  1. Independence of non-overlapping time intervals
  2. Customers won’t cluster in a short interval
  3. Probability of arrival in a short interval = length of interval

\[ F(t) = 1 - e^{-\lambda t} \]

\[ f(t) = \lambda e^{-\lambda t} \]

\[ T \sim \text{Exp}(\lambda) \]

Exponential Distribution

Example: Average 2 failures of a machine in one day

- Assume Poisson process
  - time unit = day
  - \( \lambda = 2 \)

\[ T = \text{time till the first failure} \]

\[ \lambda = 2 \]

\[ E(T) = \frac{1}{\lambda} \]

\[ \text{Var}(T) = \frac{1}{\lambda^2} \]

<table>
<thead>
<tr>
<th>Time</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.135335</td>
</tr>
<tr>
<td>1</td>
<td>0.270670</td>
</tr>
<tr>
<td>2</td>
<td>0.270670</td>
</tr>
<tr>
<td>3</td>
<td>0.205005</td>
</tr>
<tr>
<td>4</td>
<td>0.141903</td>
</tr>
<tr>
<td>5</td>
<td>0.091268</td>
</tr>
<tr>
<td>6</td>
<td>0.060845</td>
</tr>
<tr>
<td>7</td>
<td>0.040589</td>
</tr>
<tr>
<td>8</td>
<td>0.028412</td>
</tr>
<tr>
<td>9</td>
<td>0.019608</td>
</tr>
<tr>
<td>10</td>
<td>0.013072</td>
</tr>
</tbody>
</table>

\[ Pr(T > 4) = 1 - F(4) = 1 - (1 - e^{-2}) = 0.000335 \]

\[ Pr(T > a + b \mid T > a) = Pr(T > b) \]

Memoryless
Gamma Distribution

Gamma function: $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$

$\Gamma(1) = 1$

$\Gamma(\alpha) = (\alpha-1) \Gamma(\alpha - 1)$

$\Gamma(\alpha) = (\alpha - 1)!$

Gamma distribution: $X \sim \Gamma(\alpha, \lambda)$

$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x > 0$

$E(X) = \frac{\alpha}{\lambda}$

$Var(X) = \frac{\alpha}{\lambda^2}$

Example: Average 20 phone calls per hour

Assume Poisson process

Assume $\lambda = 20$

Time unit = hour

$T = \text{waiting time until the 8th call}$

$E(T) = \frac{8}{20} = 0.4 \text{ hour} = 24 \text{ minutes}$

$Pr(T > 0.8) = 1 - Pr(T \leq 0.8) = 1 - \int_0^{0.8} \frac{20}{\Gamma(8)} x^{7} e^{-20x} dx = \text{?}$

Chi-Square Distribution

$X \sim \Gamma(\alpha, \lambda) \iff aX \sim \Gamma\left(\frac{\alpha}{a}, \frac{\lambda}{a}\right)$

$X \sim \Gamma(\alpha, \lambda) \iff 2\lambda X \sim X^2_{2\alpha}$

Example: $T = \Gamma(8,20) \iff 40T \sim \Gamma\left(\frac{8}{2}, \frac{1}{2}\right) = X^2_{16}$

$Pr(T \leq 0.8) = Pr(40T \leq 32) = Pr(X^2_{16} \leq 32) = 0.99$

Chi-Square Distribution Table

$X = \chi^2$

$Pr(X \leq 15.51) = 0.95$
Normal Distribution
\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty \]
\[ N(\mu, \sigma^2) \]

Standard Normal Distribution
\[ X \sim N(0,1) \]
\[ \Phi(x) = \Pr(X \leq x) = \Phi(x) = 0.9798 \text{ for } c = 2.05 \]
\[ \Phi(0.917) = 0.8204 \]

Standard Normal Distribution Table

Normal Distribution
\[ aX + b = N(\mu + b, a^2 \sigma^2) \]
\[ X \sim N(\mu, \sigma^2) \]
\[ \Phi(x) = \Pr(X \leq x) = \Phi(x) = 0.9798 \]
\[ \Phi(0.917) = 0.8204 \]

Normal Distribution Example
Overall scores approximately normal with mean 74.3 and sd 11.8
\[ X \sim N(74.3, 11.8) \]
\[ \Pr(X \geq 85) = 1 - \Phi \left( \frac{85 - 74.3}{11.8} \right) = 1 - \Phi(0.907) = 1 - 0.8178 = 0.1822 \]
\[ 95\% \text{ can pass} \Rightarrow \text{passing mark} = 7 \]