Probability and Statistics

November 14, 2018

Contents

1 Generic Formulas

1.1 Expected Value

X is discrete random variable
$g : \mathbf{R} \to \mathbf{R}$
$\Omega X = Im(X)$

$$E[g(X)] = \sum_{x \in \Omega X} P(X = x) * g(x) \tag{1}$$

1.2 Variance

$$\begin{aligned}
var(g(X)) &= E([g(X) - E(g(X))]^2) && (2)\\
&= \sum_{x \in \Omega X} [g(X) - E(g(X))]^2 * P(X = x)] && (3)\\
&= \sum_{x \in \Omega X} [g(X)^2 - 2 * g(X) * E(g(X)) + E(g(X))^2] * P(X = x) && (4)\\
&= \sum_{x \in \Omega X} g(X)^2 * P(X = x) && (5)\\
&- \sum_{x \in \Omega X} 2 * g(X) * E(g(X)) * P(X = x) && (6)\\
&+ \sum_{x \in \Omega X} E(g(X))^2 * P(X = x) && (7)\\
&= \sum_{x \in \Omega X} g(X)^2 * P(X = x) && (8)\\
&- 2 * E(g(X)) * \sum_{x \in \Omega X} g(X) * P(X = x) && (9)\\
&+ E(g(X))^2 * \sum_{x \in \Omega X} P(X = x) && (10)\\
&= E(g(X)^2) - 2 * E(g(X)) * E(g(X)) + E(g(X))^2 * 1 && (11)\\
&= E(g(X)^2) - 2 * E(g(X))^2 + E(g(X))^2 && (12)\\
&= E(g(X)^2) - E(g(X))^2 && (13)\\
\square & && (14)
\end{aligned}$$

1.3 Covariance Matrix

$$
\begin{aligned}
cov(X) &= E[(X - E(X)]^2 && (15)\\
&= \sum_{x \in \Omega X} (X - E(X))^2 * P(X = x) && (16)\\
&= \sum_{x \in \Omega X} [X^2 - 2XE(X) + E(X)^2] * P(X = x) && (17)\\
&= \sum_{x \in \Omega X} X^2 P(X = x) && (18)\\
&- \sum_{x \in \Omega X} 2XE(X) * P(X = x) && (19)\\
&+ \sum_{x \in \Omega X} E(X)^2 P(X = x) && (20)\\
&= \sum_{x \in \Omega X} X^2 P(X = x) && (21)\\
&- 2E(X) * \sum_{x \in \Omega X} XP(X = x) && (22)\\
&+ E(X)^2 * \sum_{x \in \Omega X} P(X = x) && (23)\\
&= E(X^2) - 2E(X)E(X) + E(X)^2 * 1 && (24)\\
&= E(X^2) - 2 * E(X)^2 + E(X)^2 && (25)\\
&= E(X^2) - E(X)^2 && (26)\\
&= E(X^t X) - \mu^t \mu && (27)\\
\square & && (28)
\end{aligned}
$$

1.4 Variance of the Sample Mean

$$Var\left(\overline{X}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \tag{29}$$

$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \tag{30}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i), \text{ by independence} \tag{31}$$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + \ldots + Var(X_n)\right] \tag{32}$$

$$= \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \ldots + \sigma^2\right], \text{ since the } X_i \text{ are identically distributed} \tag{33}$$

$$= \frac{1}{n^2}(n\sigma^2) \tag{34}$$

$$= \frac{\sigma^2}{n} \tag{35}$$

1.5 Law of Iterated Expectation

$$E[X] = E[E[X|Y]] \tag{36}$$

1.6 Law of Total Variance

$$var(X) = E[var(X|Y)] + var(E[X|Y]) \tag{37}$$

1.7 MSE

$$\hat{\theta} = \hat{\theta}(X) \text{ Random Variable} \tag{38}$$
$$E[\hat{\theta}] = \text{constant} \tag{39}$$
$$\theta = \text{true value, constant} \tag{40}$$
$$\tag{41}$$
$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{42}$$
$$= E[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}]) - \theta)^2] \tag{43}$$
$$= E[([\hat{\theta} - E[\hat{\theta}]] + [E[\hat{\theta}] - \theta])^2] \tag{44}$$
$$= E[(A + B)^2] \tag{45}$$
$$\tag{46}$$
$$A = \hat{\theta} - E[\hat{\theta}] \tag{47}$$
$$B = E[\hat{\theta}] - \theta \tag{48}$$
$$\tag{49}$$
$$= E[A^2 + 2AB + B^2] \tag{50}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2 + 2(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta) + (E[\hat{\theta}] - \theta)^2] \tag{51}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + E[2(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)] + E[(E[\hat{\theta}] - \theta)^2] \tag{52}$$
$$\tag{53}$$
$$C = E[\hat{\theta}] - \theta \text{ is a constant} \tag{54}$$
$$\tag{55}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + E[2(\hat{\theta} - E[\hat{\theta}])(C)] + E[C^2] \tag{56}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2CE[\hat{\theta} - E[\hat{\theta}]] + C^2 \tag{57}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2C(E[\hat{\theta}] - E[E[\hat{\theta}]]) + C^2 \tag{58}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2C(E[\hat{\theta}] - E[\hat{\theta}]) + C^2 \tag{59}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2C(0) + C^2 \tag{60}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + C^2 \tag{61}$$
$$= E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta)^2 \tag{62}$$
$$\tag{63}$$
$$var(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2] \tag{64}$$
$$bias(\hat{\theta}, \theta) = E[\hat{\theta}] - \theta \tag{65}$$
$$\tag{66}$$
$$= var(\hat{\theta}) + (bias(\hat{\theta}, \theta))^2 \tag{67}$$
$$\tag{68}$$

2 Binomial Distribution

PMF

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{69}$$

2.1 Expected Value

$E[g(X)]$ when $g(X) = X$.

$$\begin{aligned} E(X) &= \sum_{k \geqslant 0} P(x = k) * k & (70) \\ &= \sum_{k \geqslant 0} [\binom{n}{k} p^k (1-p)^{n-k}] * k & (71) \\ & & (72) \end{aligned}$$

when

$$k = 0$$

, the formula

$$[\binom{n}{k} p^k (1-p)^{n-k}] * k = [\binom{n}{0} p^k (1-p)^n] * 0 = 0$$

, so the index of the summation can be increased by 1.

$$
\begin{aligned}
E(X) &= \sum_{k\geqslant 1} \binom{n}{k} p^k (1-p)^{n-k} * k && (73)\\
&= \sum_{k\geqslant 1} \frac{n}{k} * \binom{n-1}{k-1} p^k (1-p)^{n-k} * k \qquad \text{see BinomialCoefficient} && (74)\\
&= \sum_{k\geqslant 1} \frac{n*k}{k} * \binom{n-1}{k-1} p^k (1-p)^{n-k} && (75)\\
&= \sum_{k\geqslant 1} n * \binom{n-1}{k-1} p^k (1-p)^{n-k} && (76)\\
&= \sum_{k\geqslant 1} n*p* \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} && (77)\\
&= np * \sum_{k\geqslant 1} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} && (78)\\
u &= n-1 && (79)\\
z &= k-1 && (80)\\
u-z &= (n-1)-(k-1) && (81)\\
&= n-1-k+1 && (82)\\
&= n-k && (83)\\
k>1 &= (z+1)>1 && (84)\\
&= z>0 && (85)\\
&= np * \sum_{z>0} \binom{u}{z} p^z (1-p)^{u-z} && (86)\\
&= np*1 \qquad see Binomial Distribution Proof Equals 1 && (87)\\
&= np && (88)\\
&\square && (89)
\end{aligned}
$$

2.2 Variance

$$
\begin{aligned}
Var(X) &= E(X^2) - E(X)^2 \qquad \text{see Variance} && (90)\\
&= \sum_{k\geqslant 0} [\binom{n}{k} p^k (1-p)^{n-k}] * k^2 - np \qquad \text{see Binomial Expected Value} && (91)
\end{aligned}
$$

when

$$k = 0$$

, the formula

$$[\binom{n}{k}p^k(1-p)^{n-k}]*k = [\binom{n}{0}p^k(1-p)^n]*0 = 0$$

, so the index of the summation can be increased by 1.

$$= \sum_{k\geqslant 1} [\binom{n}{k} p^k (1-p)^{n-k}] * k^2 - (np)^2 \tag{92}$$

$$= \sum_{k\geqslant 1} \frac{n}{k} [\binom{n-1}{k-1} p^k (1-p)^{n-k}] * k^2 - (np)^2 \tag{93}$$

$$= \sum_{k\geqslant 1} \frac{n * k^2}{k} [\binom{n-1}{k-1} p^k (1-p)^{n-k}] - (np)^2 \tag{94}$$

$$= \sum_{k\geqslant 1} [nk * \binom{n-1}{k-1} p^k (1-p)^{n-k}] - (np)^2 \tag{95}$$

$$= \sum_{k\geqslant 1} [nkp * \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}] - (np)^2 \tag{96}$$

$$= np * \sum_{k\geqslant 1} [k * \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}] - (np)^2 \tag{97}$$

$$u = n - 1 \tag{98}$$

$$z = k - 1 \tag{99}$$

$$u - z = (n-1) - (k-1) \tag{100}$$

$$= n - 1 - k + 1 \tag{101}$$

$$= n - k \tag{102}$$

$$k >= 1 = (z+1) >= 1 \tag{103}$$

$$= z >= 0 \tag{104}$$

$$= np * \sum_{z\geqslant 0} [(z+1) * \binom{u}{z} p^z (1-p)^{u-z}] - (np)^2 \tag{105}$$

$$= np * [\sum_{z\geqslant 0} [z * \binom{u}{z} p^z (1-p)^{u-z}] + \sum_{z\geqslant 0} [\binom{u}{z} p^z (1-p)^{u-z}]] - (np)^2 \tag{106}$$

$$= np * [\sum_{z\geqslant 0} [z * \frac{u}{z} * \binom{u-1}{z-1} p^z (1-p)^{u-z}] + \sum_{z\geqslant 0} [\binom{u}{z} p^z (1-p)^{u-z}]] - (np)^2 \tag{107}$$

$$= np * [u * \sum_{z\geqslant 0} [\binom{u-1}{z-1} p^z (1-p)^{u-z}] + \sum_{z\geqslant 0} [\binom{u}{z} p^z (1-p)^{u-z}]] - (np)^2 \tag{108}$$

$$\tag{109}$$

$$= np * [up * \sum_{z \geqslant 0} [\binom{u-1}{z-1} p^{z-1} (1-p)^{u-z}] + \sum_{z \geqslant 0} [\binom{u}{z} p^z (1-p)^{u-z}]] - (np)^2 \tag{110}$$

$$= np * [up * \sum_{z \geqslant 1} [\binom{u-1}{z-1} p^{z-1} (1-p)^{(u-1)-(z-1)}] + \sum_{z \geqslant 0} [\binom{u}{z} p^z (1-p)^{u-z}]] - (np)^2 \tag{111}$$

$$= np * [up * \sum_{z \geqslant 0} [\binom{u}{z} p^z (1-p)^{u-z}] + \sum_{z \geqslant 0} [\binom{u}{z} p^z (1-p)^{u-z}]] - (np)^2 \tag{112}$$

$$= np * [up * (p+q)^{u-1} + (p+q)^u] - (np)^2 \tag{113}$$

$$= np * [(n-1) * p * (p+q)^{n-1-1} + (p+q)^(n-1)] - (np)^2 \tag{114}$$

$$= np * [(n-1) * p * (p+q)^{n-2} + (p+q)^(n-1)] - (np)^2 \tag{115}$$

$$= np * ([(n-1) * p * (p+q)^{n-2} + (p+q)^(n-1)] - np) \tag{116}$$

$$= np * ([(n-1) * p + 1] - np) \qquad \text{p+q=1} \tag{117}$$

$$= np * ([(n-1) * p + 1] - np) \tag{118}$$

$$= np * ([np - p + 1] - np) \tag{119}$$

$$= np * (np - p + 1 - np) \tag{120}$$

$$= np * (-p + 1) \tag{121}$$

$$= np * (1-p) \tag{122}$$

$$\square \tag{123}$$

3 Bernoulli Distribution

The Bernoulli Distribution is a special case of the Binomial Distribution, where

$$n = 1$$

.

PMF

$$P(X = k) = \binom{1}{k} p^k (1-p)^{1-k} \tag{124}$$

$$= p^k (1-p)^{n-k} \tag{125}$$

3.1 Expected Value

$$E(x) = \sum_{k\geqslant 1}[\binom{n}{k}p^k(1-p)^{n-k}] * k \tag{126}$$
$$= np \qquad seeBinomialDistribution \tag{127}$$
$$= 1 * p \tag{128}$$
$$= p \tag{129}$$

3.2 Variance

$$Var(X) = np * (1-p) \tag{130}$$
$$= p * (1-p) \tag{131}$$
$$\tag{132}$$

3.3 Likelihood of IID Bernoulli

$$x_i \overset{iid}{\sim} Bernoulli(p) \tag{133}$$
$$L(x_i|p) = p(x_1, x_2, ..., x_n|p) \tag{134}$$
$$= \prod_{n=1}^{n} p(x_i|p) \tag{135}$$
$$= p^S * (1-p)^{n-S} \tag{136}$$

3.4 Maximun Likelihood

$$\begin{aligned}
\frac{d[L(x_i|p)]}{dp} &= \frac{d[p^S * (1-p)^{n-S}]}{dp} && (137)\\
\frac{d[log(L(x_i|p))]}{dp} &= \frac{d[log(p^S * (1-p)^{n-S})]}{dp} && (138)\\
&= \frac{d}{dp} * [log(p^S * (1-p)^{n-S})] && (139)\\
&= \frac{d}{dp}[log(p^S) + log((1-p)^{n-S})] && (140)\\
&= \frac{d}{dp}[S * log(p) + (n-S) * log(1-p)] && (141)\\
&= S * \frac{d}{dp}[log(p)] + (n-S) * \frac{d}{dp}[log(1-p)] && (142)\\
&= S * [\frac{1}{p}] + (n-S) * \frac{d}{dp}[log(1-p)] \qquad \text{chain rule} && (143)\\
&= S * \frac{1}{p} + (n-S) * \frac{1}{p-1} && (144)\\
&= \frac{S}{p} + \frac{n-S}{p-1} && (145)\\
&= \frac{S*(p-1)}{p*(p-1)} + \frac{p*(n-S)}{p*(p-1)} && (146)\\
&= \frac{S*(p-1) + p*(n-S)}{p*(p-1)} && (147)\\
&= \frac{S*p - S + p*n - p*S}{p*(p-1)} && (148)\\
&= \frac{-S + p*n}{p*(p-1)} && (149)\\
0 &= \frac{-S + p*n}{p*(p-1)} && (150)\\
0*(p*(p-1)) &= -S + p*n && (151)\\
0 &= -S + p*n && (152)\\
S &= p*n && (153)\\
\frac{S}{n} &= p && (154)\\
p &= \frac{S}{n} && (155)\\
& && (156)
\end{aligned}$$

4 Geometric Distribution

4.1 CDF

$$
\begin{aligned}
P(X \leq x) &= \\
CDF(X = x) &= \sum_{i=0}^{x} (1-p)^i p && \text{by geometric summation} \\
&= p \frac{1-(1-p)^x}{1-(1-p)} \\
&= p \frac{1-(1-p)^x}{1-1+p} \\
&= p \frac{1-(1-p)^x}{p} \\
&= 1-(1-p)^x \\
\square
\end{aligned}
$$

$$
\begin{aligned}
P(X > x) &= 1 - CDF(X = x) \\
&= 1 - (1 - (1-p)^x) \\
&= 1 - 1 + (1-p)^x \\
&= (1-p)^x \\
\square
\end{aligned}
$$

5 Normal Distribution

5.1 Definition

$$pdf(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{157}$$

Continuous Distributions

6 Uniform Distribution

6.1 PDF

$$\begin{aligned}\int_a^b k dx &= 1 \\ &= k\int_a^b dx \\ &= k[x|_a^b] \\ &= k[b-a]\end{aligned}$$

$$\begin{aligned}k[b-a] &= 1 \\ k &= \frac{1}{b-a}\end{aligned}$$

6.2 Expected Value

$$\begin{aligned}\int_a^b x(\frac{1}{b-a})dx &= \\ &= \int_a^b x(\frac{1}{b-a})dx \\ &= \frac{1}{b-a}\int_a^b xdx \\ &= \frac{1}{b-a}[\frac{x^2}{2}|_a^b] \\ &= \frac{1}{b-a}[\frac{b^2}{2}-\frac{a^2}{2}] \\ &= \frac{1}{b-a}[\frac{b^2-a^2}{2}] \\ &= \frac{1}{b-a}[\frac{(b+a)(b-a)}{2}] \\ &= \frac{(b+a)(b-a)}{2(b-a)} \\ &= \frac{b+a}{2}\end{aligned}$$

6.3 Variance

$$
\begin{aligned}
var(X) &= E[X^2] - E[X]^2 \\
&= [\int_a^b x^2 \Big(\frac{1}{b-a}\Big) dx] - (\frac{b+a}{2})^2 \\
&= [\frac{1}{b-a} \int_a^b x^2 dx] - (\frac{b+a}{2})^2 \\
&= [\frac{1}{b-a} \frac{x^3}{3} |_a^b] - (\frac{b+a}{2})^2 \\
&= \frac{1}{b-a} (\frac{b^3}{3} - \frac{a^3}{3}) - (\frac{b+a}{2})^2 \\
&= \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} \\
&= \frac{(b-a)(b^2 - ab + a^2)}{3(b-a)} - \frac{(b+a)^2}{4} \\
&= \frac{(b^2 - ab + a^2)}{3} - \frac{(b+a)^2}{4} \\
&= \frac{4(b^2 - ab + a^2)}{12} - \frac{3(b+a)^2}{12} \\
&= \frac{4b^2 - 4ab + 4a^2}{12} - \frac{3(b^2 + 2ab + a^2)}{12} \\
&= \frac{4b^2 - 4ab + 4a^2}{12} - \frac{3b^2 + 6ab + 3a^2}{12} \\
&= \frac{4b^2 - 3b^2 - 4ab - 6ab + 4a^2 - 3a^2}{12} \\
&= \frac{b^2 - 2ab + a^2}{12} \\
&= \frac{(b-a)^2}{12}
\end{aligned}
$$

7 Exponential Distribution

7.1 PDF

$$f_x(x) = -\lambda e^{-\lambda x} \qquad\qquad x >= 0$$

7.2 Expected Value

$$f_x(x) = \lambda e^{-\lambda x}$$

$$E[x] = \int x * f_x(x) dx$$
$$E[x] = \int x * \lambda e^{-\lambda x} dx$$

$$u = x$$
$$du = 1 * dx$$
$$v = e^{-\lambda x}$$
$$dv = e^{-\lambda x} * -\lambda * dx$$
$$dv = -\lambda e^{-\lambda x} dx$$

$$E[x] = \int [x][\lambda e^{-\lambda x} dx]$$
$$E[x] = -1 * \int [x][-1 * \lambda e^{-\lambda x} dx]$$
$$E[x] = -1 * \int [x][-\lambda e^{-\lambda x} dx]$$
$$E[x] = -1 * \int udv$$
$$E[x] = -1 * [uv|_{-\infty}^{\infty} - \int vdu]$$
$$E[x] = -1 * [uv|_{0}^{\infty} - \int vdu] \qquad \text{because x bounds}$$
$$E[x] = -1 * [[xe^{-\lambda x}]_{0}^{\infty} - \int e^{-\lambda x} dx]$$

$$
\begin{aligned}
[xe^{-\lambda x}]_0^{\infty} &= \lim_{x->\infty} [xe^{-\lambda x} - (0)e^{-\lambda(0)}] \\
&= \lim_{x->\infty} [xe^{-\lambda x} + 0e^{0}] \\
&= \lim_{x->\infty} [xe^{-\lambda x}] \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
E[x] &= -1 * - \int e^{-\lambda x} dx \\
E[x] &= \frac{1}{\lambda} \int \lambda e^{-\lambda x} dx \\
E[x] &= \frac{1}{\lambda}
\end{aligned}
$$