Set Theory

Let P be a partition of X, then
$\bigcup P = X,\ A \cap B = \emptyset$ if $A \in P, B \in P, A \neq B$

Discrete Mathematics

$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \binom{n}{k} = \binom{n}{n-k}$ for $0 \leq k \leq n$

Recursive formula: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$

Pascal's rule: $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

Calculus

$\frac{d}{dx} f(x)g(x) = f'(x)g(x) + f(x)g'(x)$

$\frac{P(x)}{Q(x)} = \frac{c_1}{x-a_1} + \frac{c_2}{x-a_2} + \cdots + \frac{c_n}{x-a_n}$

If $deg\,P \leq deg\,Q \Rightarrow P(x) \div Q(x) = D(x) + \frac{R(x)}{Q(x)}$

$\frac{P(x)}{Q(x)} = \frac{P(x)}{(x-a)^r} = \frac{c_1}{x-a} + \frac{c_2}{(x-a)^2} + \cdots + \frac{c_r}{(x-a)^r}$

Probability

$P(\emptyset) = 0, P(\Omega) = 1, P(A^c) = 1 - P(A), A^c \cup A = \Omega$

$P(A^c) + P(A) = P(A^c \cup A) = P(\Omega) = 1$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) +$
$P(A \cap B) = P(A \cup B) + P(A \cap B)$

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

A, B independent if $P(A \cap B) = P(A) \cdot P(B)$

Bayes: $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

Total Probability Law:
$P(B) = \sum_i [P(B|A_i) \cdot P(A_i)]$

Random Variables

pdf disc: $f(x) = P(X = x)$, *cdf*: $F(x) = P(X \leq x)$
$\sum_{x \in \Omega} f(x) = 1, \quad \int f(x) = 1$

pdf cont: $P(a \leq X \leq x) = \int_a^b f(x)dx$

Expectation

$\mathrm{E}[h(X)] = \sum_{all\ x} h(x) f(x)$

$\mathrm{E}[X] = \int x \cdot f(x)$

$\mathrm{E}[a] = a, \mathrm{E}[aX] = a\mathrm{E}[X]$

$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

Variance

$Var(X) = \sum_{all\ x} f(x)(x - \mu)^2$

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

$Var(a) = 0, \quad Var(aX) = a^2 Var(X)$

$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

$E\left[\frac{X-\mu}{\sigma}\right] = 0, Var\left(\frac{X-\mu}{\sigma}\right) = 1$

Continuous Random Variables

$P[X = a] = P[X = b] = 0 \Rightarrow P[a \leq X \leq b] =$
$P[a < X < b], f(x) = F'(x)$

Multivariate Random Variables

$f_{XY}(x, y) = P[X = x \text{ and } Y = y]$

$f_X(x) = \sum_{all\ y} f_{XY}(x, y)$

$f_Y(y) = \sum_{all\ x} f_{XY}(x, y)$

$Cov(X, Y) = \sigma_{XY} =$
$\mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])] =$
$\mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]$

$Cov(X, a) = 0, \quad Cov(X, X) = Var(x)$

$Cov(X, Y) = Cov(Y, X)$

$Cov(aX, bY) = abCov(X, Y)$

$Cov(X + a, Y + b) = Cov(X, Y)$

$Corr(X, Y) = \varphi_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$

$X = Y \Rightarrow Corr(X, Y) = 1$

Independent Random Variables

$\mathrm{E}[XY] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$

$Var(X + Y) = Var(X) + Var(Y)$

independent $\Rightarrow Cov(X, Y) = 0$

$Cov(X, Y) = 0 \Leftrightarrow Corr(X, Y) = 0$

Distributions

Geometric Distribution

$X \sim Geom(p)$

$f(x) = (1 - p)^{x-1} p$

$F(x) = 1 - (1 - p)^{\lfloor x \rfloor}$

$\mathrm{E}[X] = \frac{1}{p}, Var(X) = \frac{1-p}{p^2}$

Binomial Distribution

Fixed number, n, of Bernoulli trials, independent trials with equal probability p. $X \sim Bin(n, p)$

$f(x) = \binom{n}{k} p^k (1 - p)^{n-k}$

$\mathrm{E}[X] = np, Var(X) = np(1 - p)$

Continuous Uniform Distribution

$X \sim U(a, b)$

$f(x) = \frac{1}{b-a}, F(x) = \frac{x-a}{b-a}$

$\mathrm{E}[X] = \frac{1}{2}(a + b), Var(X) = \frac{1}{12}(b - a)^2$

Normal Distribution

$X \sim N(\mu, \sigma^2)$

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$

Standard Normal Distribution

$X \sim N(0, 1)$

Poisson Distribution

$X \sim Poisson(\lambda), f(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$

$\mathrm{E}[X] = \lambda, \mathrm{E}\left[X^2\right] = \lambda^2 + \lambda, Var(X) = \lambda$

Exponential Distribution

$X \sim Exp(\lambda)$

$f(x) = \lambda e^{-\lambda x}, F(x) = 1 - e^{-\lambda x}$

$\mathrm{E}[X] = \frac{1}{\lambda}, \mathrm{E}\left[X^2\right] = \frac{2}{\lambda^2}, Var(X) = \frac{1}{\lambda^2}$

Estimation

$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

$\hat{\theta}$ is unbiased if $\mathrm{E}\left[\hat{\theta}\right] = \theta$

$\widehat{p_1 - p_2} = \widehat{p_1} - \widehat{p_2} = \frac{X_1}{n} - \frac{X_2}{n}$

Central Limit Theorem

If $|\mathrm{E}[X]| < \infty$ and $Var(X) < \infty$ then
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow \frac{\bar{X}-\mu}{\frac{\sigma^2}{n}} \sim N(0, 1)$

Confidence Interval

$Z \sim N(0, 1), T \sim t(n - 1), C \sim \chi^2(n - 1)$

$\mu \to \bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$\mu \to \bar{X} \pm T_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$

$\sigma^2 \to \left[\frac{(n-1)S^2}{C_{\frac{\alpha}{2}}}, \frac{(n-1)S^2}{C_{1-\frac{\alpha}{2}}}\right]$

$p \to \hat{p} \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}$

$p_1 - p_2 \to \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

For one sided confidence intervals, use Z_α

Markov Chains

$P(X_{n+1} = i \,|\, X_n = a_n, X_{n-1} = a_{n-1}, \ldots, X_0 =$
$a_0) = P(X_{n+1} = i \,|\, X_n = a_n)$

$p_{ij}^{(n)} = P(X_n = j \,|\, X_0 = i)$

$p_{ij} = P(X_1 = j \,|\, X_0 = i)$

Absorption

State absorbing if $p_{ii} = 1$ and $p_{ij} = 0$ for $i \neq j$

j is the starting state, a is the absorbing state, the probability q_j is

$$q_j = \begin{cases} 1 & j = a \\ 0 & j \neq a \text{ and } j \text{ absorbing} \\ \sum_k P_{jk} \cdot q_k & \text{otherwise} \end{cases}$$

Number of steps to absorption, m_j, is

$$m_j = \begin{cases} 0 & j \text{ absorbing} \\ \sum_k P_{jk} \cdot m_k & \text{otherwise} \end{cases}$$

Hypothesis Testing

Decide H_0, H_0 True $\Rightarrow$ OK
Decide H_0, H_A True $\Rightarrow$ Type II error
Decide H_A, H_0 True $\Rightarrow$ Type I error
Decide H_A, H_A True $\Rightarrow$ OK

Test statistics

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}, t = \frac{\bar{x}-\mu}{S/\sqrt{n}}, f = \frac{s_1^2}{s_2^2}$

Generating Functions

$G(a_n; x) = A(x) = \sum_{n=0}^\infty a_n x^n$

$\frac{\mathrm{d}^n A}{\mathrm{d}x^n}\Big|_{x=0} = a_n \cdot n! \Rightarrow a_n = \frac{A^{(n)}(0)}{n!}$

Common Operations

$A(x) + B(x) \to \{a_n + b_n\}_{n=0}^\infty$

$xA(x) + a_{-1} \to a_{-1}, a_0, a_1, \ldots$

$\frac{A(x)-a_0}{x} \to a_1, a_2, a_3, \ldots$

$A'(x) \to \{n \cdot a_n\}_{n=0}^\infty$

$A(x) \cdot B(x) \to c_n =$
$a_0 b_n + a_1 b_{n-1} + \cdots + a_{n-1} b_1 + a_n b_0$

Moment Generating Functions

$\mathrm{E}[X^n]$ is the n-th moment

$\mathrm{E}[X - \mathrm{E}[X]]^n$ is the n-th central moment

$m_X(t) = \mathrm{E}\left[e^{tX}\right]$

$m_X^{(n)}(0) = \mathrm{E}[X^n]$

$m_X(t) = m_Y(t) \Rightarrow$ X, Y equally distributed

Let X, Y be independent with m.g.f. $m_X(t), m_Y(t)$ then $Z = X + Y$ has m.g.f. $m_Z(t) = m_X(t) \cdot m_Y(t)$

Inequalities

Markov's Inequality

$X > 0$ and $t > 0 \Rightarrow P(X \geq t) \leq \frac{\mathrm{E}[X]}{t}$

Chebyshev's Inequality

$\varepsilon > 0, P(|X - \mu| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$

Law of Large Numbers

Let $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$, and $|\mathrm{E}[X]| < \infty$, $Var(X) < \infty$, then
$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1$

Sequences

Geometric

$x^0 + x^1 + \cdots + x^k = \sum_{n=0}^k x^n = \frac{1-x^{k+1}}{1-x}$

$c^0 x^0 + c^1 x^1 + c^2 x^2 + \cdots = \sum_{k=0}^\infty (cx)^k = \frac{1}{1-cx}$

$x^k + x^{k+1} + x^{k+2} + \cdots = \sum_{i=0}^\infty x^{k+i} = \frac{x^k}{1-x}$

$x^{0k} + x^{1k} + x^{2k} + \cdots = \sum_{n=0}^\infty x^{nk} = \frac{1}{1-x^k}$

Binomial

$\binom{n}{0}x^0 + \binom{n}{1}x^1 + \binom{n}{2}x^2 + \cdots = \sum_{k=0}^\infty \binom{n}{k} x^k = (1+x)^n$

$\binom{k}{k}x^0 + \binom{k+1}{k}x^1 + \binom{k+2}{k}x^2 + \cdots =$
$\sum_{n=0}^\infty \binom{n+k}{k} x^n = \frac{1}{(1-x)^{k+1}}$

$\sum_{k=0}^n \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = (x + y)^n$

MacLaurin

$\frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^\infty \frac{x^n}{n!} = e^x$

Moment Generating Functions

$m_X(t) = \sum_{n=0}^{\infty} \left[\frac{\mathrm{E}[X]^n}{n!} t^n\right] = \mathrm{E}\left[\sum_{n=0}^{\infty} \left[\frac{(Xt)^n}{n!}\right]\right] = \mathrm{E}\left[e^{tx}\right]$

Let X have m.g.f. $m_X(t)$ then $Y = a + bX$ has m.g.f: $m_Y(t) = \mathrm{E}\left[e^{tY}\right] = \mathrm{E}\left[e^{t(a+bX)}\right] = \mathrm{E}\left[e^{at}e^{btX}\right] = e^{at} \cdot \mathrm{E}\left[e^{btX}\right] = e^{at} \cdot m_X(bt)$

Let X, Y be independent with m.g.f. $m_X(t)$, $m_Y(t)$, then $Z = X + Y$ has m.g.f: $m_Z(t) = \mathrm{E}\left[e^{tZ}\right] = \mathrm{E}\left[e^{t(X+Y)}\right] = \mathrm{E}\left[e^{tX}e^{tY}\right] = \mathrm{E}\left[e^{tX}\right] \cdot \mathrm{E}\left[e^{tY}\right] = m_X(t) \cdot m_Y(t)$

Geometric Distribution

$m_X(t) = \sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1}p = pe^t \sum_{x=1}^{\infty} e^{t(x-1)}(1-p)^{x-1} = pe^t \sum_{x=0}^{\infty} e^{tx}(1-p)^x = pe^t \sum_{x=0}^{\infty} (e^t(1-p))^x = \frac{pe^t}{1-e^t(1-p)}$

Binomial Distribution

$m_X(t) = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x(1-p)^{n-x} = (pe^t + (1-p))^n = (p(e^t-1)+1)^n$

Exponential Distribution

Only if $t < \lambda$

$m_X(t) = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \lambda\int_0^{\infty} e^{x(t-\lambda)}dx = \lambda\left[\frac{e^{x(t-\lambda)}}{t-\lambda}\right]_0^{\infty} = \lambda\left(0 - \frac{1}{t-\lambda}\right) = \frac{\lambda}{\lambda - t}$

$F_X(x) = P(X \le x) = \int_0^x f_X(s)ds = \int_0^x \lambda e^{-\lambda s}ds = \lambda\left[-\frac{1}{\lambda}e^{-\lambda s}\right]_0^x = (-1)(e^{-\lambda x} - 1) = 1 - e^{-\lambda x}$

Poisson Distribution

$m_X(t) = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(e^t\lambda)^x}{x!} = e^{\lambda(e^t-1)}$

Inequalities

Markov

$\mathrm{E}[X] = \int_0^{\infty} xf(x)\,dx = \int_0^t xf(x)\,dx + \int_t^{\infty} xf(x)\,dx \ge 0 + \int_t^{\infty} tf(x)\,dx = t\int_t^{\infty} f(x)\,dx = tP(X \ge t) \Rightarrow \frac{E[X]}{t} \ge P(X \ge t)$

Chebyshev

Use Markov's inequality: $P(|X-\mu| \ge \varepsilon) = P(|X-\mu|^2 \ge \varepsilon^2) \le \frac{\mathrm{E}[|X-\mu|]^2}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$

Law of Large Numbers

Use Markov's inequality to prove this theorem.

Confidence Intervals

$X_1, \ldots, X_n; X_i \sim N(\mu, \sigma^2)$

Use the CLT to get $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

We can then construct a $1-\alpha$ level CI interval: $P(-Z_{\frac{\alpha}{2}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le Z_{\frac{\alpha}{2}}) = P(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le Z_{\frac{\alpha}{2}}) = P(|\bar{X}-\mu| \le Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = P(\bar{X} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$

Markov Chains

If X_n has distribution u then X_{n+1} has distribution uP:

$P(X_{n+1} = j) = \sum_i [P(X_n = i) \cdot P(X_{n+1} = j \mid X_n = i)] = \sum_i [u_i P_{ij}] = (uP)_j$

Absorbing Chain

Assume j not absorbing and let A = {chain absorbed in a}: $q_j = P(A \mid X_0 = j) = \frac{P(A \cap (X_0 = j))}{P(X_0 = j)} = \frac{\sum_k P(A \cap (X_0 = j) \cap (X_1 = k))}{P(X_0 = j)} = \sum_k \left[P(A \mid X_0 = j \cap X_1 = k) \cdot \frac{P(X_0 = j \cap X_1 = k)}{P(X_0 = j)}\right] = \sum_k [P(A \mid X_1 = k) \cdot P(X_1 = k \mid X_0 = j)] = \sum_k [P(A \mid X_1 = k) \cdot P_{jk}] = \sum_k q_k P_{ij}$

Expectation

Function of X

$\mathrm{E}[h(X)] = \sum_k \left[k \cdot f_{h(X)}(k)\right] = \sum_k [k \cdot P(h(X) = k)] = \sum_k \left[k \cdot \sum_{x\in\Omega, h(X)=k} [P(X = x)]\right] = \sum_{x\in\Omega}\left[\sum_{k,h(X)=k} [k \cdot P(X = x)]\right] = \sum_{x\in\Omega} [h(x) \cdot P(X = x)] = \sum_{x\in\Omega} [h(x) \cdot f_x(x)]$

Multiplication

If X, Y independent

$\mathrm{E}[XY] = \sum_{x,y} [xy \cdot f_{XY}(x,y)] = \sum_x \left[\sum_y [xy \cdot f_X(x)f_Y(y)]\right] = \sum_x [xf_X(x)] \cdot \sum_y [yf_Y(y)] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$

Addition

$\mathrm{E}[X+Y] = \sum_r [r \cdot P(X + Y = r)] = \sum_{j,k} [(x_j + y_k)P(X = x_j, Y = y_k)] = \sum_{j,k} [x_j \cdot P(X = x_j, Y = y_k)] + \sum_{j,k} [y_k \cdot P(X = x_j, Y = y_k)] = \sum_j \left[x_j \sum_k [P(X = x_j, Y = y_k)]\right] + \sum_k \left[y_k \sum_j [P(X = x_j, Y = y_k)]\right] = \sum_j [x_j \cdot P(X = x_j)] + \sum_k [y_k \cdot P(Y = y_k)] = \mathrm{E}[X] + \mathrm{E}[Y]$

Variance

Just X

$Var(X) = \mathrm{E}[X - \mathrm{E}[X]]^2 = \mathrm{E}\left[X^2 - 2X\mathrm{E}[X] + \mathrm{E}[X]^2\right] = \mathrm{E}\left[X^2\right] - 2\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}\left[\mathrm{E}[X]^2\right] = \mathrm{E}\left[X^2\right] - 2\mathrm{E}[X]^2 + \mathrm{E}[X]^2 = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2$

Addition

$Var(X+Y) = \mathrm{E}\left[(X+Y)^2\right] - \mathrm{E}[X+Y]^2 = \mathrm{E}\left[X^2 + Y^2 + 2XY\right] - (\mathrm{E}[X] + \mathrm{E}[Y])^2 = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 + \mathrm{E}[Y] - \mathrm{E}[Y]^2 + 2\mathrm{E}[XY] - 2\mathrm{E}[X]\mathrm{E}[Y] = Var(X) + Var(Y) + 2Cov(X,Y)$

Covariance

$Cov(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[XY - X\mathrm{E}[Y] - Y\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[Y]] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] - \mathrm{E}[Y]\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$

Other

Bayes

$P(A|B) = \frac{P(A\cap B)}{P(B)}$ *and* $P(B|A) = \frac{P(A\cap B)}{P(A)} \Rightarrow P(A\cap B) = P(A|B)P(B) = P(B|A)P(A) \Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Marginal distribution

$f_X(x) = P(X = x) = P(\bigcup_y \{X = x, Y = y\}) = \sum_y [P(X = x, Y = y)] = \sum_y [f_{XY}(x,y)]$

Bernoulli

$Y = 1$ if successes otherwise 0 with probability p and $(1-p)$

$\mathrm{E}[Y] = p, \mathrm{E}\left[Y^2\right] = p$

$Var(Y) = \mathrm{E}\left[Y^2\right] - \mathrm{E}[Y]^2 = p - p^2 = p(1-p)$

$X = \sum_{i=1}^n Y_i, X \sim Bin(n,p) \Rightarrow \mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n \mathrm{E}[Y_i] = \sum_{i=1}^n p = np$

$Var(X) = Var(\sum_{i=1}^n Var(Y_i)) = \sum_{i=1}^n p(1-p) = np(1-p)$

Estimators

Unbiased

$\hat{\theta}$ is unbiased if $\mathrm{E}\left[\hat{\theta}\right] = \theta, \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \mathrm{E}\left[\bar{X}\right] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i] = \mu$

Variance

Proof for the computational formula:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}\left[X_i - \bar{X}\right]^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - 2X_i\bar{X} + \bar{X}^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i\bar{X} + \sum_{i=1}^{n}\bar{X}^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{2}{n}\left[\sum_{i=1}^{n} X_i\right]^2 + \frac{n}{n^2}\left[\sum_{i=1}^{n} X_i\right]^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{2}{n}\left[\sum_{i=1}^{n} X_i\right]^2 + \frac{1}{n}\left[\sum_{i=1}^{n} X_i\right]^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left[\sum_{i=i}^{n} X_i\right]^2\right]$$

Note that S^2 is also unbiased while S is biased, because math.