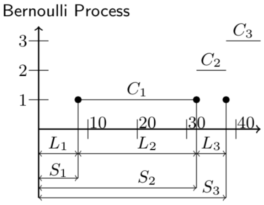
Chapter 2

- Random Variables
  - Function on Ω, generalizes an event
- Probability Mass Function
  - $P_X(i) = P(X = i)$
  - Sum of PMF $= 1$
- Expected Value (average)
  - $E[X] = \mu_X = \Sigma u_i \cdot P_x(u_i)$
  - LOTUS
    - $E[Y] = E[g(X)] = \Sigma g(u_i) \cdot P_X(u_i)$
  - Scaling
    - $E[aX] = a \cdot E[X]$
  - Addition
    - $E[X+b] = b + E[X]$
- Poisson Distribution
  - $X \sim \text{Pois}(\lambda)$
  - $E[X] = \lambda$
  - $\text{Var}(X) = \lambda$
  - $P(X = k) = \dfrac{\lambda^k \cdot e^{-\lambda}}{k!}$
  - $X \sim \text{Bin}(n,p) \approx \text{Pois}(np), n \gg p$
- Negative Binomial Distribution
  - Sum of Geometric
  - $S_r \sim \text{NB}(r,p)$
  - $P(S_r = n) = p^r \dbinom{n-1}{r-1} \cdot (1-p)^{n-r}$
  - $E[S_r] = \dfrac{r}{p}$
  - $\text{Var}(S_r) = \dfrac{r(1-p)}{p^2}$
- Maximum Likelihood Estimation
  - Take Derivative, Set $= 0$ (can use ln)
  - observe $X = k$, p unknown

    $Bin(n,p)$ $\quad \widehat{p}_{ML} = \dfrac{k}{n}$ $\qquad Geo(n)$ $\quad \widehat{p}_{ML} = \dfrac{1}{k}$ $\qquad Pois(\lambda)$ $\quad \widehat{\lambda}_{ML}(k) = k$
- Markov Inequality
  - $P(Y \geq c) \leq \dfrac{E[Y]}{c} = \dfrac{\mu_Y}{c}, \quad Y \geq 0, c > 0$
- Chebychev Inequality
  - $P(|Y - E[Y]| \geq c) \leq \dfrac{\sigma_Y^2}{c^2}$
  - Confidence Interval $\quad X \sim Bin(n,p)$
    - $P\left(|X - n \cdot p| \geq \dfrac{a \cdot \sqrt{n}}{2}\right) \leq \dfrac{1}{a^2}$
    - $P\left(\left|\dfrac{X}{n} - p\right| < \dfrac{a}{2\sqrt{n}}\right) \geq 1 - \dfrac{1}{a^2}$
    - $P\left(p \in \left[\widehat{p} - \dfrac{a}{2\sqrt{n}}, \widehat{p} + \dfrac{a}{2\sqrt{n}}\right]\right) \geq 1 - \dfrac{1}{a^2}, \quad \widehat{p} = \dfrac{X}{n}$
- Law of Total Probability
  - $P(A) = \sum_{i=1}^{n} P(E_i) \cdot P(A|E_i), \quad \bigcup_{i=1}^{n} E_i = \Omega$
- Bayes Formula
  - $P(B|A) = \dfrac{P(BA)}{P(A)} = \dfrac{P(B) \cdot P(A|B)}{P(A)}$
- Independence
  - $A \perp\!\!\!\perp B \iff P(AB) = P(A) \cdot P(B)$
  - $A, B, C \perp\!\!\!\perp \iff A, B, C$ pairwise independent and $P(ABC) = P(A) \cdot P(B) \cdot P(C)$
  - Given $X \perp\!\!\!\perp Y$
    - $f(X) \perp\!\!\!\perp g(Y)$
    - $E[XY] = E[X] \cdot E[Y]$
    - $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
- Bernoulli Distribution
  - $X \sim \text{Ber}(p) = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$
  - $E[X] = p$
  - $\text{Var}(X) = p \cdot (1-p)$

- Hypothesis testing
  - $P_{\text{false alarm}} = P(\text{declare } H_1 | H_0)$
  - $P_{\text{miss}} = P(\text{declare } H_0 | H_1)$
  - $P_e = P(H_0) \cdot P_{\text{false alarm}} + P(H_1) \cdot P_{\text{miss}}$
  - $\Lambda(j) = \dfrac{P(X = j | H_1)}{P(X = j | H_0)}$
  - Maximum Likelihood
    - $\Lambda(X) \underset{H_0}{\overset{H_1}{\gtrless}} 1$
  - Maximum a Posteriori Probability
    - $\Lambda(X) \underset{H_0}{\overset{H_1}{\gtrless}} \dfrac{P(H_0)}{P(H_1)} = \dfrac{\pi_0}{\pi_1}$
- Bernoulli Process



- Binomial Distribution
  - Sum of n Bernoulli w.p. p
  - $S \sim \text{Bin}(n,p) = \sum_{i=1}^{n} X_i$
  - $E[S] = n \cdot p$
  - $\text{Var}(S) = n \cdot p \cdot (1-p)$
  - $P(S = k) = \dbinom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$
- Geometric Distribution
  - $Z \sim \text{Geo}(p)$
  - $E[Z] = \dfrac{1}{p}$
  - $\text{Var}(Z) = \dfrac{1-p}{p^2}$
  - $P(Z = k) = p(1-p)^{k-1}$

Erlang(r, λ)**:** $\quad r \geq 1, \lambda \geq 0$

pdf: $f(t) = \dfrac{\lambda^r t^{r-1} e^{-\lambda t}}{(r-1)!} \quad t \geq 0 \quad$ mean: $\dfrac{r}{\lambda} \quad$ variance: $\dfrac{r}{\lambda^2}$

Significant property: The distribution of the sum of r independent random variables, each having the exponential distribution with parameter λ.

- Union Bound
  - $P(A \cup E$
  - Good wh
  - $P(A_1 \cup$
- Network Reli
  - Use Law
  - Use Union Bound for Accurate and Easy Answer
- Array Code
  - Error Detection

$$p_S(k) = P\{X + Y = k\} = \sum_j P\{X = j, Y = k - j\} = \sum_j p_{X,Y}(j, k-j).$$

$$\begin{aligned} MSE &= E[(Y - \mu_Y - a(X - \mu_X))^2] \\ &= \text{Var}(Y - aX) \\ &= \text{Cov}(Y - aX, Y - aX) \\ &= \text{Var}(Y) - 2a\text{Cov}(Y, X) + a^2\text{Var}(X). \end{aligned}$$

$$\begin{aligned} p_S(k) &= \sum_{j=0}^{\min(m,k)} \binom{m}{j} p^j (1-p)^{m-j} \binom{n}{k-j} p^{k-j} (1-p)^{n-k+j} \\ &= \left(\sum_{j=0}^{\min(m,k)} \binom{m}{j} \binom{n}{k-j} \right) p^k (1-p)^{m+n-k} \\ &= \binom{m+n}{k} p^k (1-p)^{m+n-k}, \end{aligned}$$

$$\begin{aligned} p_{X+Y}(k) &= \sum_{j=0}^{k} \frac{\lambda_1^j e^{-\lambda_1}}{j!} \frac{\lambda_2^{k-j} e^{-\lambda_2}}{(k-j)!} \\ &= \left(\sum_{j=0}^{k} \frac{\lambda_1^j \lambda_2^{k-j}}{j!(k-j)!} \right) e^{-(\lambda_1 + \lambda_2)} \\ &= \left(\sum_{j=0}^{k} \binom{k}{j} p^j (1-p)^{k-j} \right) \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \frac{\lambda^k e^{-\lambda}}{k!}, \end{aligned}$$

$$\begin{aligned} f_S(c) &= \frac{dF_S(c)}{dc} = \int_{-\infty}^{\infty} \frac{d}{dc} \left(\int_{-\infty}^{c-u} f_{X,Y}(u,v)dv \right) du \\ &= \int_{-\infty}^{\infty} f_{X,Y}(u, c-u)du. \end{aligned}$$

$$f_S = f_X * f_Y \qquad \text{(if } S = X + Y\text{, where } X, Y \text{ are independent).}$$

$$f_{W,Z}(\alpha, \beta) = \frac{1}{|\det A|} f_{X,Y}\left(A^{-1}\begin{pmatrix}\alpha \\ \beta\end{pmatrix}\right)$$

the correlation: $E[XY]$

the covariance: $\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$

the correlation coefficient: $\rho_{X,Y} = \dfrac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \dfrac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}.$$

Furthermore, if $E[X^2] \neq 0$, equality holds in (4.26) (i.e. $|E[XY]| = \sqrt{E[X^2]E[Y^2]}$) if and only if $P\{Y = cX\} = 1$ for some constant c.

$$\text{MSE (for estimation of } Y \text{ by a constant } \delta) = \int_{-\infty}^{\infty} (y - \delta)^2 f_Y(y)dy$$

It remains to find the constant a. The MSE is quadratic in a, so taking the derivative with respect to a and setting it equal to zero yields that the optimal choice of a is $a^* = \frac{\text{Cov}(Y,X)}{\text{Var}(X)}$. Therefore, the minimum MSE linear estimator is given by $L^*(X) = \widehat{E}[Y|X]$, where

$$\widehat{E}[Y|X] = \mu_Y + \left(\frac{\text{Cov}(Y,X)}{\text{Var}(X)}\right)(X - \mu_X) \tag{4.35}$$

$$= \mu_Y + \sigma_Y \rho_{X,Y} \left(\frac{X - \mu_X}{\sigma_X}\right). \tag{4.36}$$

Setting a in (4.34) to a^* gives the following expression for the minimum possible MSE:

$$\text{minimum MSE for linear estimation} = \sigma_Y^2 - \frac{(\text{Cov}(X,Y))^2}{\text{Var}(X)} = \sigma_Y^2(1 - \rho_{X,Y}^2). \tag{4.37}$$

$$\text{minimum MSE for linear estimation} = \sigma_Y^2 - \text{Var}(\widehat{E}[Y|X]) = E[Y^2] - E[\widehat{E}[Y|$$

$$\underbrace{E[(Y - g^*(X))^2]}_{\text{MSE for } g^*(X) = E[Y|X]} \leq \underbrace{\sigma_Y^2(1 - \rho_{X,Y}^2)}_{\text{MSE for } L^*(X) = \widehat{E}[Y|X]} \leq \underbrace{\sigma_Y^2}_{\text{MSE for } \delta^* = E[Y].}$$

$$P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \delta\right\} \leq \frac{C}{n\delta^2} \xrightarrow{n \to \infty} 0.$$

$$\begin{aligned}\mathrm{Var}(S_n) &= \mathrm{Cov}(X_1+\cdots+X_n, X_1+\cdots+X_n)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i,X_j)\\ &= \sum_{i=1}^{n-1}\mathrm{Cov}(X_i,X_{i+1}) + \sum_{i=1}^{n}\mathrm{Cov}(X_i,X_i) + \sum_{i=2}^{n}\mathrm{Cov}(X_i,X_{i-1})\\ &\le (n-1)(0.1)+n+(n-1)(0.1) < (1.2)n.\end{aligned}$$

$$\lim_{n\to\infty} P\left\{\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le c\right\} = \Phi(c).$$

$$\begin{aligned}\rho_{X,Y} = E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv f_X(u) f_{Y|X}(v|u)\,dv\,du\\ &= \int_{-\infty}^{\infty} u f_X(u)\left(\int_{-\infty}^{\infty} v f_{Y|X}(v|u)\,dv\right)du\\ &= \rho\int_{-\infty}^{\infty} u^2 f_X(u)\,du = \rho.\end{aligned}$$

Exponential(λ): $\lambda > 0$

pdf: $f(t) = \lambda e^{-\lambda t}$ $t \ge 0$ mean: $\frac{1}{\lambda}$ variance: $\frac{1}{\lambda^2}$

Example: Time elapsed between noon sharp and the first time a telephone call is placed after that, in a city, on a given day.

Significant property: If T has the exponential distribution with parameter λ, $P\{T \ge t\} = e^{-\lambda t}$ for $t \ge 0$. So T has the *memoryless property* in continuous time:

$$P\{T \ge s+t | T \ge s\} = P\{T \ge t\} \qquad s,t \ge 0$$

Any nonnegative random variable with the memoryless property in continuous time is exponentially distributed. Failure rate function is constant: $h(t) \equiv \lambda$.

Rayleigh(σ^2): $\sigma^2 > 0$

pdf: $f(r) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)$ $r > 0$ CDF : $1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)$

mean: $\sigma\sqrt{\frac{\pi}{2}}$ variance: $\sigma^2\left(2 - \frac{\pi}{2}\right)$

Example: Instantaneous value of the envelope of a mean zero, narrow band noise signal.

Significant property: If X and Y are independent, $N(0,\sigma^2)$ random variables, then $(X^2+Y^2)^{\frac{1}{2}}$ has the Rayleigh(σ^2) distribution. Failure rate function is linear: $h(t) = \frac{t}{\sigma^2}$.

$$E[Y|X=10] = \int_{-\infty}^{\infty} v f_{Y|X}(v|10)\,dv.$$

The resulting conditional MSE is the variance of Y, computed using the conditional distribution of Y given $X = 10$.

$$E[(Y - E[Y|X=10])^2|X=10] = E[Y^2|X=10] - (E[Y|X=10]^2).$$

Conditional expectation indeed gives the optimal estimator, as we show now. Recall that $f_{X,Y}(u,v) = f_X(u) f_{Y|X}(v|u)$. So

$$\begin{aligned}\mathrm{MSE} &= E[(Y-g(X))^2]\\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}(v-g(u))^2 f_{Y|X}(v|u)\,dv\right) f_X(u)\,du. \qquad (4.29)\end{aligned}$$

For each u fixed, the integral in parentheses in (4.29) has the same form as the integral (4.28). Therefore, for each u, the integral in parentheses in (4.29) is minimized by using $g(u) = g^*(u)$, where

$$g^*(u) = E[Y|X=u] = \int_{-\infty}^{\infty} v f_{Y|X}(v|u)\,dv. \qquad (4.30)$$

We write $E[Y|X]$ for $g^*(X)$. The minimum MSE is

$$\begin{aligned}MSE &= E[(Y - E[Y|X])^2]\\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}(v-g^*(u))^2 f_{Y|X}(v|u)\,dv\right) f_X(u)\,du \qquad (4.31)\\ &\overset{a}{=} \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}(v^2-(g^*(u))^2) f_{Y|X}(v|u)\,dv\right) f_X(u)\,du \qquad (4.32)\\ &= E[Y^2] - E[(E[Y|X])^2], \qquad (4.33)\end{aligned}$$

where the equality (a) follows from the shortcut $\mathrm{Var}(Y) = E[Y^2] - E[Y]^2$, applied using the conditional distribution of Y given $X = u$. In summary, the minimum MSE unconstrained estimator of Y given X is $E[Y|X] = g^*(X)$ where $g^*(u) = E[Y|X=u]$, and expressions for the MSE are given by (4.31)-(4.33).

$$f_{X,Y}(u,v) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{\left(\frac{u-\mu_X}{\sigma_X}\right)^2 + \left(\frac{v-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{u-\mu_X}{\sigma_X}\right)\left(\frac{v-\mu_Y}{\sigma_Y}\right)}{2(1-\rho^2)}\right)$$

$$f_{W,Z}(\alpha,\beta) = \left(\frac{e^{-\frac{\alpha^2}{2}}}{\sqrt{2\pi}}\right)\left(\frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}}\right) = \frac{e^{-\frac{\alpha^2+\beta^2}{2}}}{2\pi}$$

(a) X has the $N(\mu_X,\sigma_X^2)$ distribution, and Y has the $N(\mu_Y,\sigma_Y^2)$ distribution.

(b) Any linear combination of the form $aX + bY$ is a Gaussian random variable (i.e., X and Y are jointly Gaussian).

(c) ρ is the correlation coefficient between X and Y (i.e. $\rho_{X,Y} = \rho$).

(d) X and Y are independent if and only if $\rho = 0$.

(e) For estimation of Y from X, $L^*(X) = g^*(X)$. Equivalently, $E[Y|X] = \widehat{E}[Y|X]$. That is, the best unconstrained estimator $g^*(X)$ is linear.

(f) The conditional distribution of Y given $X = u$ is $N(\widehat{E}[Y|X=u],\sigma_e^2)$, where σ_e^2 is the MSE for $\widehat{E}[Y|X]$, given by (4.37) or (4.38).

$$\begin{aligned}f_{X,Y}(u,v) &= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right)\\ &= \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)\right]\left[\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right)\right]\\ f_X(u) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)\\ f_{Y|X}(v|u) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right)\end{aligned}$$

Example 4.11.4 Let X and Y be jointly Gaussian random variables with mean zero, variance one, and $\mathrm{Cov}(X,Y) = \rho$. Find $E[Y^2|X]$, the best estimator of Y^2 given X. (Hint: X and Y^2 are not jointly Gaussian. But you know the conditional distribution of Y given $X = u$ and can use it to find the conditional second moment of Y given $X = u$.)

Solution: Recall the fact that $E[Z^2] = E[Z]^2 + \mathrm{Var}(Z)$ for a random variable Z. The idea is to apply the fact to the conditional distribution of Y given X. Given $X = u$, the conditional distribution of Y is Gaussian with mean ρu and variance $1-\rho^2$. Thus, $E[Y^2|X=u] = (\rho u)^2 + 1 - \rho^2$. Therefore, $E[Y^2|X] = (\rho X)^2 + 1 - \rho^2$.

Geometric(p): $0 < p \le 1$

pmf: $p(i) = (1-p)^{i-1}p$ $i \ge 1$

mean: $\frac{1}{p}$ variance: $\frac{1-p}{p^2}$

Example: Number of independent flips of a coin until heads first shows.

Significant property: If L has the geometric distribution with parameter p, $P\{L > i\} = (1-p)^i$ for integers $i \ge 1$. So L has the *memoryless property* in discrete time:

$$P\{L > i+j | L > i\} = P\{L > j\} \text{ for } i,j \ge 0.$$

Any positive integer-valued random variable with this property has the geometric distribution for some p.