Expectation of Functions of RVs and Laws of Large Numbers

- Expectation of Functions of RVs
- Moments
- Moment generation function
- Characteristic function
- Inequalities/Bounds
- Laws of large numbers
- Central limit theorem
- Entropy

Expectation of Functions of RVs

- For Continuous RVs:

$$Y = g(X) \;\Rightarrow\; \boxed{E[Y] = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_{-\infty}^{\infty} g(x) f_X(x)dx}$$

$$Z = g(X,Y) \Rightarrow \boxed{E[Z] = \int_{-\infty}^{\infty} z f_Z(z)dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)dxdy}$$

- Discrete RVs:

$$Y = g(X) \;\Rightarrow\; \boxed{E[Y] = \sum_k y_k P_Y(y_k) = \sum_i g(x_i) P_X(x_i)}$$

$$Z = g(X,Y) \;\Rightarrow\; \boxed{E[Z] = \sum_m P_Z(z_m) = \sum_k \sum_i g(x_i, y_k)\; P_{XY}(x_i, y_k)}$$

- Linearity of Expectation: $\boxed{E\left[\sum_{i=1}^N g_i(X)\right] = \sum_{i=1}^N E[g_i(X)]}$

e.g., $E[X^3 - 5X + \frac{1}{X} - 2] = E[X^3] - 5E[X] + E[\frac{1}{X}] - 2$

Conditional Expectation

- Continuous: $\boxed{E[X|B] = \int_{-\infty}^{\infty} x \ f_{X|B}(x|B) \ dx}$
- Discrete: $\boxed{E[X|B] = \sum_i x_i \ P_{X|B}(x_i|B)}$
- Discrete $X\&Y$: $\boxed{E[Y|X = x_i] \triangleq \sum_j y_j P_{Y|X}(y_j|x_i)}$
- $\boxed{E[Y] = \sum_i E[Y|X = x_i] P_X(x_i)}$
- Continuous $X\&Y$: $\boxed{E[Y|X = x] \triangleq \int_{-\infty}^{\infty} y \ f_{Y|X}(y|x) \ dy}$
- $\boxed{E[Y] = \int_{-\infty}^{\infty} E[Y|X = x] \ f_X(x) \ dx}$
- Continuous X, Discrete Y: $\boxed{E[Y|X = x] = \sum_j y_j \ P_{Y|X}(y_j|x)}$
- Discrete X, Continuous Y: $\boxed{E[Y|X = x_i] = \int y \ f_{Y|X}(y|x_i) \ dy}$

- Conditional Expectation as a RV (when the conditioned parameters are RVs):

  $E[Y|X]$ is a function of X, say $g(X)$, and hence a RV.

$$\boxed{\begin{aligned} E[Y] &= E[E[Y|X]] \qquad \text{(Note: Outer } E[\cdot] \text{ is } E_X[\cdot]) \\ &= \begin{cases} \sum_i E[Y|X=x_i]\; P_X(x_i), & \text{discrete} \\ \int_{-\infty}^{\infty} E[Y|X=x] f_X(x)\; dx, & \text{continuous} \end{cases} \end{aligned}}$$

$$\boxed{\begin{aligned} E[Z] &= E[E[Z|X,Y]] \qquad \text{(Note: Outer } E[\cdot] \text{ is } E_{X,Y}[\cdot]) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z\; f_{Z|X,Y}(z|x,y)\; f_{XY}(x,y)\; dx\; dy\; dz \end{aligned}}$$

$$\boxed{E[Z|X] = E[E[Z|X,Y]|X] \qquad \text{(Note: Outer } E[\cdot] \text{ is } E_{Y|X}[\cdot] = h(X))}$$

- Law of total Variance:

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

- Example: Consider a communication system in which the message delay (in milliseconds) is T and the channel choice is L. Let $L = 1$ for a satellite channel, $L = 2$ for a coaxial cable channel, $L = 3$ for a microwave terrestrial link, and $L = 4$ for a fiber-optical link. A channel is chosen based on availability, which is a random phenomenon. Suppose $P_L(l) = 1/4$, $l = 1, 2, 3, 4$. It is known that $E[T|L = 1] = 500$, $E[T|L = 2] = 300$, $E[T|L = 3] = 200$, and $E[T|L = 4] = 100$. Find $E[T|L]$.

$$E[T|L] = g(L) = \begin{cases} 500, & L = 1 \\ 300, & L = 2 \\ 200, & L = 3 \\ 100, & L = 4 \end{cases}$$

$$E[T] = E[E[T|L]] = \sum_{l=1}^{4} E[T|L = l]P_L[l] = (500 + 300 + 200 + 100)/4 = 275$$

- In a photoelectric detector, the number of photoelectrons Y produced in time τ depends on the (normalized) incident energy X. If X were constant, say $X = x$, Y would be a Poisson RV with parameter x (=mean). However, in practice the pdf of X is well modeled by $f_X(x) = \frac{1}{\mu_X} \exp(-x/\mu_X)\, u(x)$, i.e., exponential with $E[X] = \mu_X$. Find $E[Y]$.

$P_{Y|X}[k|x] = \frac{x^k}{k!} e^{-x}, \;\; k = 0, 1, 2, \ldots$

$E[Y|X = x] = x$

$E[Y] = E[E[Y|X]] = \int_{-\infty}^{\infty} E[Y|X = x] f_X(x)dx = \int_0^{\infty} x \frac{1}{\mu_X} \exp(-x/\mu_X)\, dx = \mu_X$

(Note: Easier than computing $P_Y[k] = \int_{-\infty}^{\infty} P_{Y|X}[k|x] f_X(x)dx$ and $E[Y] = \sum_k k P_Y[k]$.)

Moments

- rth moment of X: $\boxed{m_r \triangleq E[X^r]}$, $r = 0, 1, 2, \ldots;\quad (E[X] = \mu)$
- rth central moment of X: $\boxed{c_r \triangleq E[(X-\mu)^r]}$, $\quad r = 0, 1, 2, \ldots$

$$(X-\mu)^r = \sum_{i=0}^{r} \binom{r}{i} (-1)^i \mu^i X^{r-i} \quad \Rightarrow \quad \boxed{c_r = \sum_{i=0}^{r} \binom{r}{i} (-1)^i \mu^i m^{r-i}}$$

- n^{th} absolute moment: $E[|X|^n]$
- n^{th} absolute central moment: $E[|X-\mu|^n]$
- Generalized moments: $E[(X-a)^n]$, $E[|X-a|^n]$
- Kurtosis: $\boxed{\text{Kurt}[X] = \frac{E[(X-\mu)^4]}{\sigma^4}}$; (a measure of "tailedness" or outliers)

  With $Z = (X-\mu)/\sigma$, $\text{Kurt}[X] = E[Z^4] = \text{Var}[Z^2] + 1$,
  (a measure of dispersion of Z^2 around its expectation)

  (Kurtosis of a normal RV is 3.)

- Joint Moments:

ijth joint moment of X and Y: $\boxed{m_{ij} \triangleq E[X^i Y^j]}$

ijth joint central moment of X and Y: $\boxed{c_{ij} \triangleq E[(X-\bar{X})^i (Y-\bar{Y})^j]}$

(The order of moment is $i+j$).

Second-order moments are:

$$m_{02} = E[Y^2], \qquad m_{20} = E[X^2], \qquad m_{11} = E[XY],$$

$$c_{02} = E[(Y-\bar{Y})^2], \qquad c_{20} = E[(X-\bar{X})^2],$$

$$c_{11} = E[(X-\bar{X})(Y-\bar{Y})] = E[XY] - \bar{X}\bar{Y} \triangleq \mathrm{Cov}[X,Y].$$

- Correlation coefficient:

$$\boxed{\begin{array}{l} \rho \triangleq \dfrac{c_{11}}{\sqrt{c_{20}c_{02}}} = \dfrac{\mathrm{Cov}[X,Y]}{\sigma_X \sigma_Y}, \quad |\rho| \leq 1 \\ \rho = 0 \Rightarrow \text{uncorrelated} \Rightarrow E[XY] = E[X]E[Y], \;\; \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y \\ E[XY] = 0 \Rightarrow \text{orthogonal} \end{array}}$$

Moment Generating Function (MGF)

- MGF of X, if it exists, is

$$\boxed{\theta(t) \triangleq E[e^{tX}]}, \quad \text{where } t \text{ is a complex variable}$$

$$\text{Continuous}: \theta(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

$$\text{Discrete}: \theta(t) = \sum_i e^{tx_i} P_X(x_i)$$

Except sign reversal in the exponent, MGF is two-sided Laplace transform of pdf.

- Second Moment Function (Cumulant Generating Function): $\Psi_2(t) \triangleq \ln \theta(t)$
- Cumulants: $\lambda_n \triangleq \left[\frac{d^n \Psi_2(t)}{dt^n}\right]_{t=0}$

  $\lambda_0 = 0, \quad \lambda_1 = \mu, \quad \lambda_2 = \sigma^2, \quad \lambda_3 = E[(X-\mu)^3], \quad \lambda_4 = E[(X-\mu)^4] - 3\sigma^4$

  For a Gaussian RV, $\lambda_n = 0$ for $n \geq 3$. (The only RV with this property).

  For independent X and Y, $\quad \lambda_{X+Y,n} = \lambda_{X,n} + \lambda_{Y,n}$.

Applications of MGF

- Convenient computation of moments:

$$\boxed{m_k = \theta^{(k)}(0),\ k = 0, 1, \ldots, \text{where } \theta^{(k)}(0) \triangleq \left[\frac{d^k}{dt^k}(\theta(t))\right]_{t=0}},$$

$$\text{because } \theta(t) = E[e^{tX}] = E\left[1 + tX + \frac{(tX)^2}{2!} + \ldots + \frac{(tX)^n}{n!} + \ldots\right]$$
$$= 1 + t\mu + \frac{t^2}{2!}m_2 + \ldots, + \frac{t^n}{n!}m_n + \ldots$$

- For $Z = \sum_{i=1}^{N} X_i$ with independent $\{X_i\}$, $\boxed{\theta_Z(t) = \prod_{i=1}^{N} \theta_{X_i}(t)}$
  $\Rightarrow\ f_Z(z) = f_{X_1}(z) * f_{X_2}(z) * \ldots * f_{X_N}(z),\quad (* = \text{convolution})$
- Estimating $f_X(x)$ from experimental measurements of the moments
- Solving problems involving the sums of RVs
- Demonstrating the Central Limit Theorem

Joint MGF

- Joint MGF of X and Y:

$$\boxed{\theta_{XY}(t_1, t_2) \triangleq E[e^{(t_1 X + t_2 Y)}]} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{t_1^i t_2^j}{i! j!} m_{ij} \quad \text{(power series expansion)}$$

Note: $\theta_{X,Y}(t_1, 0) = \theta_X(t_1)$.

- Computing joint moments of X and Y from joint MGF:

$$\boxed{m_{ln} = \theta_{XY}^{(l,n)}(0,0) \quad \text{where} \quad \theta_{XY}^{(l,n)}(0,0) \triangleq \left[\frac{\partial^{l+n} \theta_{XY}(t_1, t_2)}{\partial t_1^l \; \partial t_2^n} \right]_{t_1 = t_2 = 0}}$$

$$\begin{aligned} \text{Examples}: \theta_{XY}^{(1,0)}(0,0) &= E[X], \qquad \theta_{XY}^{(0,1)}(0,0) = E[Y], \\ \theta_{XY}^{(2,0)}(0,0) &= E[X^2], \qquad \theta_{XY}^{(0,2)}(0,0) = E[Y^2], \\ \theta_{XY}^{(1,1)}(0,0) &= E[XY] = \text{Cov}[X,Y] + E[X]E[Y] \end{aligned}$$

- Joint MGF of $X_1, X_2, \ldots, X_N$:

$$\begin{aligned} \theta_{X_1 \ldots X_N}(t_1, \ldots, t_N) &= E\left[\exp\left(\sum_{i=1}^{N} t_i X_i\right)\right] \\ &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \frac{t_1^{k_1}}{k_1!} \cdots \frac{t_N^{k_N}}{k_N!} E[X_1^{k_1} \ldots X_N^{k_N}] \end{aligned}$$

Characteristic Function

- Characteristic function (CF) of X: (replacing t of MGF with jw where $j \triangleq \sqrt{-1}$)

$$\boxed{\Phi_X(w) \triangleq E[e^{jwX}]} = \int_{-\infty}^{\infty} f_X(x) e^{jwx}\; dx$$

$$= \sum_i e^{jwx_i} P_X(x_i) \quad \text{(for discrete RV)}$$

Except for a sign difference in the exponent, it is Fourier transform of $f_X(x)$.

- Second characteristic function: $\Psi_2(w) = \ln \Phi(w)$

- For $Z = \sum_{i=1}^{N} X_i$ with independent $\{X_i\}$, $\boxed{\Phi_Z(w) = \prod_{i=1}^{N} \Phi_{X_i}(w)}$.

(Note: $f_Z(z) = f_{X_1}(z) * f_{X_2}(z) * \ldots * f_{X_N}(z)$, (* = convolution).

- Joint Characteristic Function of $X_1 X_2 \ldots X_N$:

$$\boxed{\Phi_{X_1 X_2 \ldots X_N}(w_1, \ldots, w_N) = E\left[\exp\left(j \sum_{i=1}^{N} w_i X_i\right)\right]}$$

- For $K < N$, $\Phi_{X_1 \ldots X_K}(w_1, \ldots, w_K) = \Phi_{X_1 \ldots X_N}(w_1, \ldots, w_K, 0, \ldots, 0)$.
  e.g., $\Phi_X(w_1) = \Phi_{XY}(w_1, w_2 = 0)$

- Joint pdf is the inverse Fourier transform (with a sign reversal) of the joint characteristic function:

$$f_{X_1 X_2 \ldots X_N}(x_1, \ldots, x_N) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi_{X_1 X_2 \ldots X_N}(w_1, \ldots, w_N) \times \exp\left(-j \sum_{i=1}^{N} w_i x_i\right) dw_1 \; dw_2 \ldots dw_N$$

- Computing moments from the characteristic function (CF):

$$\boxed{\begin{aligned} m_r &\triangleq E[X^r] = (-j)^r \Phi_X^{(r)}(0) = (-j)^r \Phi_{XY}^{(r,0)}(0,0), \\ m_{rk} &\triangleq E[X^r Y^k] = (-j)^{r+k} \Phi_{XY}^{(r,k)}(0,0) \\ \text{where} \quad \Phi_X^{(r)}(0) &\triangleq \left[\frac{d^r \Phi_X(w)}{dw^r}\right]_{w=0} \\ \Phi_{XY}^{(r,k)}(0,0) &\triangleq \left[\frac{\partial^{r+k} \Phi_{XY}(w_1, w_2)}{\partial w_1^r \; \partial w_2^k}\right]_{w_1 = w_2 = 0} \end{aligned}}$$

- Example: MGF and CF of $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$\theta_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx,$$

$$\text{(use the completing square approach)}$$

$$= \exp(\mu t + \frac{\sigma^2 t^2}{2}) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}) dx$$

$$= \exp(\mu t + \frac{\sigma^2 t^2}{2})$$

$$\boxed{\text{If} \quad X \sim \mathcal{N}(\mu, \sigma^2), \quad \theta_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})}$$

$\Phi_X(w) = E[e^{jwX}] = \theta_X(t = jw)$. Thus,

$$\boxed{\text{If} \quad X \sim \mathcal{N}(\mu, \sigma^2), \quad \Phi_X(w) = \exp(jw\mu - \frac{\sigma^2 w^2}{2})}$$

$$E[X] = \theta_X^{(1)}(0) = [(\mu + \sigma^2 t)\exp(\mu t + \tfrac{\sigma^2 t^2}{2})]|_{t=0} = \mu$$

$$E[X^2] = \theta_X^{(2)}(0) = [\sigma^2 \exp(\mu t + \tfrac{\sigma^2 t^2}{2}) + (\mu + \sigma^2 t)^2 \exp(\mu t + \tfrac{\sigma^2 t^2}{2})]|_{t=0} = \sigma^2 + \mu^2$$

(DIY for $E[X] = (-jw)\Phi_X^{(1)}(0)$ and $E[X^2] = (-jw)^2 \Phi_X^{(2)}(0)$)

- Example: PMF of the sum of independent Poisson RVs X and Y with parameter a and b.

$$P_X[k] = \frac{a^k}{k!}e^{-a}u[k], \quad (u[k] = \text{discrete unit step function})$$
$$P_Y[k] = \frac{b^k}{k!}e^{-b}u[k],$$

CF of X is $\Phi_X(w) = \sum_k e^{jwk}P_X[k] = \sum_{k=0}^{\infty} \frac{(a\ e^{jw})^k}{k!}e^{-a} = e^{ae^{jw}}e^{-a}$.

Similarly, CF of Poisson RV Y with parameter b is $\Phi_Y(w) = e^{b(e^{jw}-1)}$.

For $Z = X + Y$ with independent X and Y, CF of Z is

$$\Phi_Z(w) = \Phi_X(w)\Phi_Y(w) = e^{(a+b)(e^{jw}-1)}$$

which is CF of a Poisson RV with parameter $(a + b)$.

Thus,

$$P_Z[k] = \frac{(a+b)^k}{k!}e^{-(a+b)}u[k]$$

(Note: Easier than computing the convolution: $P_Z[k] = P_X[k] * P_Y[k]$)

- Example: Consider correlated bi-variate Gaussian X and Y with joint pdf

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right).$$

where $E[X] = E[Y] = 0$, $\sigma_X^2 = \sigma_Y^2 = 1$, and their correlation coefficient is ρ. Find their joint characteristic function.

Since $\sigma_X^2 = \sigma_Y^2 = 1$, $\mathrm{Cov}[X,Y] = \rho$.

Next, as $E[X] = E[Y] = 0$, $E[XY] = \mathrm{Cov}[X,Y] = \rho$.

Define $Z = w_1 X + w_2 Y$. Then, Z is $N(0, \sigma_Z^2)$ with

$$\begin{aligned} \sigma_Z^2 &= E[(\omega_1 X + \omega_2 Y)^2] = \omega_1^2 E[X^2] + \omega_2^2 E[Y^2] + 2\omega_1\omega_2 E[XY] \\ &= \omega_1^2 + \omega_2^2 + 2\omega_1\omega_2\rho \end{aligned}$$

$$\Phi_{XY}(w_1, w_2) = E[\exp(j(w_1 X + w_2 Y))] = E[\exp(jZ)] = E[\exp(jwZ)]|_{w=1}$$

$$\text{However,} \quad E[\exp(j\omega Z)] = \exp\left[-\frac{1}{2}\sigma_Z^2\omega^2\right] \quad (\text{CF of } \mathcal{N}(0,\sigma_Z^2)).$$

Hence, by setting $w = 1$,

$$\Phi_{XY}(w_1, w_2) = \exp\left(-\frac{1}{2}\left(\omega_1^2 + \omega_2^2 + \omega_1\omega_2\rho\right)\right).$$

- Example: Find the covariance of X and Y if their joint characteristic function is

$$\Phi_{XY}(w_1, w_2) = \exp\left(-\frac{1}{2}\left(\omega_1^2 + \omega_2^2 + 0.3\ \omega_1\omega_2\right)\right).$$

$$\begin{aligned}
E[XY] &= (-j)^2\ \Phi_{XY}^{(1,1)}(0,0) \\
&= (-j)^2\ \left[\frac{\partial^2 \Phi_{XY}(w_1, w_2)}{\partial w_1\ \partial w_2}\right]_{w_1=0, w_2=0} \\
&= 0.3
\end{aligned}$$

$$\begin{aligned}
\Phi_X(w_1) &= \Phi_{XY}(w_1, 0) = \exp(-\frac{w_1^2}{2}) \\
\Phi_Y(w_2) &= \Phi_{XY}(0, w_2) = \exp(-\frac{w_2^2}{2}) \\
E[X] &= (-j)\ \Phi_X^{(1)}(0) = 0 \\
E[Y] &= (-j)\ \Phi_Y^{(1)}(0) = 0
\end{aligned}$$

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 0.3$$