CS215 Expectation

Harsh Shah

August 2021

1 Introduction

Expectation of a r.v. can be considered as mean of a r.v. or equivalently the centre of mass of the probability distribution of the r.v. .
Definition(For discrete r.v.):$E[X] = \sum_x xP(x)$
Another formulation: $\sum_{s\in\Omega} X(s)P(s)$
Eg:

- For a Bernoulli r.v.(X) with parameters n, p, E[X]=np
- For a Poisson r.v.(X) with parameter λ, E[X]=λ.

In case of continuous r.v.(X), $E[X] = \int_{-\infty}^{\infty} xP(x)dx$
Another formulation: $\int_{-\infty}^{\infty} X(s)P(s)ds$
Eg:

- For exponential r.v. with average arrival rate λ(that is, $P_X(x) = \lambda e^{-\lambda x}$), $E[X] = 1/\lambda$
- For Gaussian r.v. whose distribution is given as
$$P_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$
$E[X] = \mu$

2 Linearity of expectation

If two r.v.s X and Y have the same probability space (Ω, β, P), then
$$E[X+Y] = E[X] + E[Y]$$
Proof: Definition of expected value and split the integrand of the integral.

3 Expectation of function of r.v.

Let $Y(x)$ denote a function of random variable X.
Then $E[Y] = \int_{-\infty}^{\infty} yP_Y(y)dy = \int_{-\infty}^{\infty} Y(x)P_X(x)dx = \int_{-\infty}^{\infty} Y(X(s))P(s)ds$

Important:

- $E_{P(s)}[X(s)] = E_{P(X)}[X]$
- $E_{P(Y)}[Y] = E_{P(X)}[Y(X)] = E_{P(s)}[Y(X(s))]$

Generalizing to multiple random variables $X_1 \dots X_n$:
Let $g(X_1, X_2 \dots, X_n)$ be a function of random variables, then,

$$E[g(X_1, X_2 \dots, X_n)] = \int_{-\infty}^{\infty} g(x_1, x_2 \dots, x_n) P(x_1, x_2 \dots, x_n) dx_1 dx_2 \dots dx_n$$

If X and Y are two independent r.v. :

$$E[XY] = E[X]E[Y]$$

3.1 Tail-sum formula

Let X be a discrete r.v. taking values in natural numbers. Then

$$\begin{aligned}
E[X] &= \sum_{x=1}^{\infty} x P_X(x) \\
&= \sum_{x=1}^{\infty} \sum_{k=1}^{x} P_X(x) \\
&= \sum_{k=1}^{\infty} \sum_{x=k}^{\infty} P_X(x) \\
&= \sum_{k=1}^{\infty} P_X(x \geq k)
\end{aligned}$$

If X is a continuous r.v. taking non-negative values:

$$\begin{aligned}
E[X] &= \int_0^{\infty} x f_X(x) dx \\
&= \int_0^{\infty} \int_0^{x} (f_X(x) dt) dx \\
&= \int_0^{\infty} \int_t^{\infty} f_X(x) dx dt \\
&= \int_0^{\infty} (1 - F_X(x)) dx
\end{aligned}$$

Note: $f := PDF$ and $F := CDF$

3.2 Median

Any number m is called median of a distribution $P_X(x)$ if $P_X(X \leq m) = P_X(X > m)$. Multiple medians are possible for any distribution(PDF/PMF).

3.3 Mode

For discrete X, value with maximum probability.
For continuous X, points of local maxima are called mode.

For unimodal symmetric distributions mean=median=mode.

4 Variance

Definition: For any r.v. X $var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
Standard deviation is defined as positive square root variance.
Eg:

- If X has uniform distribution on $\{a, a+1, \ldots b\}$, then $var(X) = \frac{n^2-1}{12}$ where $n = b - a + 1$
- For a binomial distribution with parameters n,p: $var(X) = np(1-p)$
- For a poisson r.v. with arrival rate λ: $var(X) = \lambda$
- For uniform distribution on (a,b): $var(X) = \frac{(b-a)^2}{12}$
- For exponential r.v. with parameter λ: $var(X) = (1/\lambda)^2$
- Gaussian r.v., with parameters (μ, σ),: $var(X) = \sigma^2$

Important properties

- $var(aX + b) = a^2 var(X)$
- $var(X + Y) = var(X) + var(Y) + 2(E[XY] - E[X]E[Y])$
  Note: If X and Y are independent then $var(X + Y) = var(X) + var(Y)$

5 Inequalities

5.1 Markov's Inequality

Consider a random variable X that always takes non-negative values. Now,

$$\begin{aligned} E[X] &= \sum_x x P_X(x) \\ &\geq \sum_{x>a} x P_X(x) \\ &\geq a P_X(X \geq a) \end{aligned}$$

In general, for any non-negative function of a r.v. (say $u(X)$),

$$P(u(X) \geq k) \leq \frac{E[u(X)]}{k}$$

5.2 Chebyshev's Inequality

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx \\ &\geq \int_{\mu+c}^{\infty} (x-\mu)^2 f_X(x) dx + \int_{-\infty}^{\mu-c} (x-\mu)^2 f_X(x) dx \\ &\geq c^2 P(|x-\mu| \geq c) \end{aligned}$$

If $c = k\sigma$,

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

5.3 Jensen's Inequality

If $f(X)$ is convex function of random variable X,

$$E[f(X)] \geq f(E[X])$$

Proof: Any tangent line lies below the function graphically.
Let the tangent line at X=E[X] be of the form $y = ax + b$
Then, $f(E[X]) = aE[X] + b$
Now,

$$E[f(X)] \geq E[aX + b] = aE[X] + b = f(E[X])$$

For concave function inequality is reversed.

5.4 Minimizer of absolute deviation

$min_m E[|X - m|]$
The solution to above optimization problem is m=median.
Proof: Split integrals about m, after breaking the problem into cases.

5.5 Theorem (mean,median,σ)

$abs(Mean - Median) \leq \sigma$
Proof:

$$abs(E[X] - median) = abs(E[X - median]) \leq E[abs(X - median)]$$
$$E[abs(X - median)] \leq E[abs(X - E[X])]$$
$$E[abs(X - E[X])]^2 \leq E[abs(X - E[X])^2] = \sigma^2$$

(QED)

5.6 Law of large numbers

Law of large numbers states that **tail probabilities** of distribution of sampling random variable

$$\widehat{X}_N = \frac{X_1 + X_2 \dots X_N}{N}$$

(where X_i are i.i.d random variables with finite mean μ and variance σ^2), tends to zero as the sample size $N \to \infty$. The proof is based on Chebyshev's inequality:

$$\begin{aligned} P(|\widehat{X}_N - E[\widehat{X}_N]| \geq \epsilon) &\leq \frac{var(\widehat{X}_N)}{\epsilon^2} \\ &\leq \frac{\sigma^2}{N\epsilon^2} \end{aligned}$$

The R.H.S. of the above inequality tends to zero as $N \to \infty$ for any $\epsilon > 0$.

6 Covariance

Definition: $E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$
Properties:

- Independence of X and Y $\implies$ cov(X,Y)=0
  Converse need not be true.
- If var(X)=0 or var(Y)=0, then cov(X,Y)=0
- If $Y = mX + c$, cov(X,Y)=$m \cdot var(X)$
- Bilinearity of covariance:cov(X+Y,Z)=cov(X,Z)+cov(Y,Z)

6.1 Standardized r.v.

For a given r.v. X, its standardized form is

$$X^* = \frac{X - E[X]}{\sigma}$$

Properties: Zero mean and unity variance.

6.2 Correlation

Definition : cor(X,Y)=$E[\frac{X-E[X]}{\sigma_X}\frac{Y-E[Y]}{\sigma_Y}] = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

Property:

- $-1 \leq cor(X,Y) \leq 1$
  Proof: $0 \leq E[(X^* + Y^*)^2]$....easy
- If $Y = mX + c$, then $|cor(X,Y)| = 1$
  Proof: Use definition of cor(X,Y)....easy
- If $|cor(X,Y)| = 1$, then X and Y are linearly dependent
  Proof: Consider different cases for $cor(X,Y) = +1, -1$,
  then use $E[(X^* - Y^*)] \geq 0$ and $E[(X^* + Y^*)] \geq 0$ to prove the claim.

Independence $\implies$ zero correlation, converse is not true.
Example: $Y = X^2$ $X \in (-1, 1)$

$$cor(X,Y) = k(E(XY) - E[X]E[Y]) = 0$$

But clearly, X and Y are dependent.

6.3 Finding line of best fit

Suppose given n samples of the form (X_i, Y_i), and that $Y = mX + c$, we intend to find the line of best fit for the given data.
We know, $E[Y] = mE[X] + c$ would hold. And
If cor(X, Y)=±1, then $Y^* = cor(X, Y) \cdot X^*$.
Simplifying the above equation

$$Y = E[Y] + \frac{cov(X, Y)(X - E[X])}{var(X)}$$

Non-Zero correlation does not imply causation.(Not concluded from above discussion)