Probability Theory Comprehensive Summary

Columbia University

December 2, 2025

Contents

1 Conditional Expectation

1.1 Conditional Expectation

Conditional Expectation

The conditional expectation of X given $Y = y$ is defined as:

- For discrete random variables:
$$E[X|Y = y] = \sum_x x \cdot g_1(x|y)$$

- For continuous random variables:
$$E[X|Y = y] = \int_{-\infty}^{\infty} x \cdot g_1(x|y) dx$$

Conditional expectation is a function of y: $h(y) = E[X|Y = y]$

Properties of Conditional Expectation

[Law of Total Expectation] For any random variables X and Y:

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

- Discrete case: $E[X] = \sum_y E[X|Y = y] \cdot f_Y(y)$
- Continuous case: $E[X] = \int_{-\infty}^{\infty} E[X|Y = y] \cdot f_Y(y) dy$

More Properties

- Linearity: $E[aX + bZ|Y] = aE[X|Y] + bE[Z|Y]$
- Taking out what's known: $E[h(Y)X|Y] = h(Y)E[X|Y]$
- Independence: If X and Y are independent, then $E[X|Y] = E[X]$
- Tower property: $E[E[X|Y, Z]|Y] = E[X|Y]$

Conditional Variance

- The conditional variance of X given $Y = y$ is:
$$\text{Var}(X|Y = y) = E[(X - E[X|Y = y])^2|Y = y]$$

- Law of Total Variance:
$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

Problem Statement

A stick of length l is broken twice:

1. **First break:** At a point X, chosen uniformly along the stick: $X \sim \text{Uniform}(0, l)$.
2. **Second break:** The smaller stick from 0 to X is taken and broken at a point Y, chosen uniformly along its length: $Y \mid X \sim \text{Uniform}(0, X)$.

Find:

1. The joint density $f(x, y)$ of X and Y.
2. The marginal density $f_Y(y)$ of Y.
3. The conditional expectation $E[Y \mid X]$ and the law of total expectation for $E[Y]$.
4. The conditional variance $\text{Var}(Y \mid X)$ and use the law of total variance to find $\text{Var}(Y)$.

Step 1: Joint Density $f(x, y)$

We use the multiplication rule for densities:

$$f(x, y) = f_X(x) \cdot g_2(y \mid x)$$

Where:

- $f_X(x)$ is the marginal density of X.
- $g_2(y \mid x)$ is the conditional density of Y given $X = x$.

From the problem:

- $X \sim \text{Uniform}(0, l) \Rightarrow f_X(x) = \frac{1}{l}$ for $0 < x < l$.
- $Y \mid X = x \sim \text{Uniform}(0, x) \Rightarrow g_2(y \mid x) = \frac{1}{x}$ for $0 < y < x$.

Therefore, the joint density is:

$$f(x, y) = \frac{1}{l} \cdot \frac{1}{x} = \frac{1}{lx}, \quad \text{for } 0 < y < x < l$$

Outside this region, the density is zero. The condition $0 < y < x < l$ ensures we are only considering the case where the first break X is to the right of the second break Y on the original stick.

Step 2: Marginal Density $f_Y(y)$

To find the marginal density of Y, we integrate the joint density over all possible values of X. For a given y, X must be greater than y (since $y < x$) and less than l.

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_{x=y}^{l} \frac{1}{lx} dx$$

Evaluating the integral:

$$f_Y(y) = \frac{1}{l} \int_y^l \frac{1}{x} dx = \frac{1}{l} [\ln x]_{x=y}^{x=l} = \frac{1}{l}(\ln l - \ln y) = \frac{1}{l} \ln\left(\frac{l}{y}\right), \quad \text{for } 0 < y < l$$

Step 3: Conditional and Total Expectation

We first find the conditional expectation $E[Y \mid X]$. Since $Y \mid X \sim \text{Uniform}(0, X)$, the mean of a uniform distribution is the midpoint:

$$E[Y \mid X] = \frac{0 + X}{2} = \frac{X}{2}$$

Now, we use the **Law of Total Expectation** to find $E[Y]$:

$$E[Y] = E[E[Y \mid X]] = E\left[\frac{X}{2}\right] = \frac{1}{2} E[X]$$

Since $X \sim \text{Uniform}(0, l)$, its expectation is $E[X] = \frac{l}{2}$.

$$E[Y] = \frac{1}{2} \cdot \frac{l}{2} = \frac{l}{4}$$

Step 4: Conditional and Total Variance

First, find the conditional variance $\text{Var}(Y \mid X)$. The variance of a Uniform$(0, X)$ distribution is $\frac{(X-0)^2}{12}$:

$$\text{Var}(Y \mid X) = \frac{X^2}{12}$$

Now, we use the **Law of Total Variance**:

$$\text{Var}(Y) = E[\text{Var}(Y \mid X)] + \text{Var}(E[Y \mid X])$$

Let's compute each term separately.

1. Expectation of the Conditional Variance:

$$E[\text{Var}(Y \mid X)] = E\left[\frac{X^2}{12}\right] = \frac{1}{12}E[X^2]$$

For $X \sim \text{Uniform}(0, l)$, $E[X^2] = \frac{l^2}{3}$.

$$E[\text{Var}(Y \mid X)] = \frac{1}{12} \cdot \frac{l^2}{3} = \frac{l^2}{36}$$

2. Variance of the Conditional Expectation:

$$\text{Var}(E[Y \mid X]) = \text{Var}\left(\frac{X}{2}\right) = \frac{1}{4}\text{Var}(X)$$

For $X \sim \text{Uniform}(0, l)$, $\text{Var}(X) = \frac{l^2}{12}$.

$$\text{Var}(E[Y \mid X]) = \frac{1}{4} \cdot \frac{l^2}{12} = \frac{l^2}{48}$$

3. Total Variance:

$$\text{Var}(Y) = \frac{l^2}{36} + \frac{l^2}{48}$$

To add these, find a common denominator, which is 144:

$$\frac{l^2}{36} = \frac{4l^2}{144}, \quad \frac{l^2}{48} = \frac{3l^2}{144}$$

$$\text{Var}(Y) = \frac{4l^2}{144} + \frac{3l^2}{144} = \frac{7l^2}{144}$$

Final Results Summary

1. **Joint Density:**

$$f(x, y) = \frac{1}{lx}, \quad \text{for } 0 < y < x < l$$

2. **Marginal Density of Y:**

$$f_Y(y) = \frac{1}{l}\ln\left(\frac{l}{y}\right), \quad \text{for } 0 < y < l$$

3. **Expectation:**

$$E[Y \mid X] = \frac{X}{2}, \quad E[Y] = \frac{l}{4}$$

4. **Variance:**

$$\text{Var}(Y \mid X) = \frac{X^2}{12}, \quad \text{Var}(Y) = \frac{7l^2}{144}$$

2 Expectations of Random Variables

2.1 Expectation of a Random Variable

- **Discrete RV:** $\mathbb{E}(X) = \sum_{\text{all } x} xf(x)$
- **Continuous RV:** $\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$
- **Bernoulli:** $\mathbb{E}(X) = p$, $\text{Var}(X) = p(1-p)$
- **Expectation of functions:** $\mathbb{E}[r(X)] = \int_{-\infty}^{\infty} r(x)f_X(x)dx$
- **Multiple RVs:** $\mathbb{E}[r(X_1, \dots, X_n)] = \int \cdots \int r(x_1, \dots, x_n)f(x_1, \dots, x_n)dx_1 \cdots dx_n$

2.2 Properties of Expectation

- **Linearity:** $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- $\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$
- **Monotonicity:** If $X \leq Y$, then $\mathbb{E}(X) \leq \mathbb{E}(Y)$
- **Jensen's Inequality:** For convex g: $\mathbb{E}[g(X)] \geq g(\mathbb{E}(X))$
- **Independence:** If X, Y independent: $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$
- **Non-negative RVs:** $\mathbb{E}(X) = \sum_{n=1}^{\infty} P(X \geq n)$ (discrete), $\mathbb{E}(X) = \int_0^{\infty} P(X > x)dx$ (continuous)

2.3 Indicator Tricks

- **Gift exchange:** $\mathbb{E}(\#$ people getting own gift$) = 1$
- **Sampling:** Expected number of red balls same with/without replacement
- **Coupon collector:** $\mathbb{E}(N) = \sum_{i=1}^{n} \frac{n}{n-i+1} = n\sum_{i=1}^{n} \frac{1}{i} \approx n \log n + \gamma n$

3 Special Distributions

3.1 Bernoulli and Binomial

- **Bernoulli:** $f(x) = p^x(1-p)^{1-x}$, $x = 0, 1$
- **Binomial:** $f(x) = \binom{n}{x}p^x(1-p)^{n-x}$, $x = 0, 1, \dots, n$
- $\mathbb{E}(X) = np$, $\text{Var}(X) = np(1-p)$, $\psi_X(t) = (pe^t + 1 - p)^n$

3.2 Hypergeometric

- $f(x) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}$
- $\mathbb{E}(X) = \frac{nA}{A+B}$, $\text{Var}(X) = \frac{nAB}{(A+B)^2}\frac{A+B-n}{A+B-1}$

3.3 Poisson Distribution

- $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2, \dots$
- $\mathbb{E}(X) = \lambda$, $\text{Var}(X) = \lambda$, $\psi_X(t) = e^{\lambda(e^t - 1)}$
- **Poisson Process:** Arrivals in time t: $\text{Pois}(\lambda t)$, independent increments

3.4 Negative Binomial and Geometric

- **Negative Binomial:** $f(x) = \binom{r+x-1}{x} p^r (1-p)^x$
- $\mathbb{E}(X) = \frac{r(1-p)}{p}$, $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$
- **Geometric:** $f(x) = p(1-p)^x$ (number of failures before first success)
- $\mathbb{E}(X) = \frac{1-p}{p}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$
- **Memoryless:** $P(X = k + t \mid X \geq k) = P(X = t)$

3.5 Normal Distribution

- $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $\mathbb{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$, $\psi_X(t) = e^{\mu t + \sigma^2 t^2/2}$
- **Standard Normal:** $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$
- **Linear combinations:** $aX + b \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$
- **Sample mean:** $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$ for i.i.d. normals

3.6 Gamma and Exponential Distributions

- **Gamma:** $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, $x > 0$
- $\mathbb{E}(X^k) = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\beta^k}$, $\mathrm{Var}(X) = \frac{\alpha}{\beta^2}$
- **Exponential:** $f(x) = \beta e^{-\beta x}$ (Gamma with $\alpha = 1$)
- $\mathbb{E}(X) = \frac{1}{\beta}$, $\mathrm{Var}(X) = \frac{1}{\beta^2}$
- **Memoryless:** $P(X > t + h \mid X > t) = P(X > h)$

3.7 Beta Distribution

- $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 \leq x \leq 1$
- $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$, $\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- $\mathbb{E}(X^k) = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{(\alpha+\beta)\cdots(\alpha+\beta+k-1)}$

3.8 Multinomial Distribution

- $f(x_1, \ldots, x_k) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k}$, $\sum x_i = n$

3.9 Bivariate Normal

- $\mathbf{X} = (X_1, X_2) \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- **Conditional:** $X_2 \mid X_1 = x_1 \sim \mathrm{N}\left(\mu_2 + \rho\sigma_2 \frac{x_1 - \mu_1}{\sigma_1}, (1-\rho^2)\sigma_2^2\right)$
- **Independence:** $\rho = 0$ iff X_1, X_2 independent

4 Large Random Samples

4.1 Law of Large Numbers

- **Markov Inequality:** $P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$ for $X \geq 0$, $a > 0$
- **Chebyshev Inequality:** $P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$
- **Weak LLN:** $P(|\bar{X}_n - \mu| \geq \epsilon) \to 0$ as $n \to \infty$
- **Convergence in probability:** $Y_n \xrightarrow{p} a$ if $\forall \epsilon > 0$, $\lim_{n\to\infty} P(|Y_n - a| \geq \epsilon) = 0$

4.2 Central Limit Theorem

- $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, 1)$
- $\bar{X}_n \approx \mathrm{N}(\mu, \sigma^2/n)$ for large n
- **Multivariate CLT:** $\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$

5 Bayesian Statistics

5.1 Prior and Posterior Distributions

- **Prior:** $f(\theta)$ - distribution before seeing data
- **Likelihood:** $f(x \mid \theta)$ - probability of data given parameter
- **Posterior:** $f(\theta \mid x) = \frac{f(x \mid \theta) f(\theta)}{\int f(x \mid \theta') f(\theta') d\theta'}$
- **Bayes Estimator:** Minimizes $\mathbb{E}[(\theta - a)^2 \mid X]$ is $\mathbb{E}[\theta \mid X]$

5.2 Conjugate Priors

Likelihood	**Prior**	**Posterior**
Bernoulli/Binomial	Beta(α, β)	Beta$(\alpha + \sum x_i, \beta + n - \sum x_i)$
Exponential	Gamma(α, β)	Gamma$(\alpha + n, \beta + \sum x_i)$

Table 1: Conjugate Prior Distributions

5.3 Bayesian Updating Examples

- **Dice identification:** Update probabilities based on rolls
- **Bent coin:** Update Beta prior with coin flip results
- **Lightbulb lifetimes:** Gamma prior with exponential likelihood
- **Sequential updating:** Posterior becomes prior for next observation

6 Important Formulas and Theorems

6.1 Distribution Relationships

- Binomial $\approx$ Poisson when n large, p small, np moderate
- Geometric is Negative Binomial with $r = 1$
- Exponential is Gamma with $\alpha = 1$
- Sample mean of i.i.d.: $\mathbb{E}(\bar{X}) = \mu$, $\text{Var}(\bar{X}) = \sigma^2/n$

6.2 Key Inequalities

- Markov: $P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$
- Chebyshev: $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$
- Jensen: $\mathbb{E}[g(X)] \geq g(\mathbb{E}(X))$ for convex g

6.3 Convergence Results

- WLLN: $\bar{X}_n \xrightarrow{p} \mu$
- CLT: $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \text{N}(0, 1)$
- Slutsky's Theorem for convergence in distribution

6.4 Bayesian Framework

- Posterior $\propto$ Likelihood $\times$ Prior
- Conjugate priors provide closed-form posteriors
- Bayes estimators minimize posterior expected loss
- Sequential updating: Process data one observation at a time

7 Applications and Examples

7.1 Probability Applications

- Coupon collector problem
- Gift exchange (derangements)
- Urn sampling with/without replacement
- Queueing theory (exponential service times)
- Stock price modeling (lognormal distribution)

7.2 Statistical Applications

- Polling and survey sampling
- Quality control and reliability
- Hypothesis testing
- Parameter estimation
- Predictive distributions

7.3 Financial Applications

- Black-Scholes option pricing
- Risk management

Appendix: Common Distributions Reference

Distribution	PMF/PDF	Mean	Variance
Bernoulli	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$
Poisson	$e^{-\lambda}\lambda^x/x!$	λ	λ
Geometric	$p(1-p)^x$	$(1-p)/p$	$(1-p)/p^2$
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Exponential	$\beta e^{-\beta x}$	$1/\beta$	$1/\beta^2$
Gamma	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$	α/β	α/β^2
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Table 2: Common Probability Distributions