

Quant 1-1

Tomoya Sasaki

July 14, 2020

Probability Theory (2, 3, 4)

Three axioms

- nonnegativity
- normalization
- additivity

↪ Check these when you see probability distributions

Key concepts

Conditional probability

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

Law of total probability

- Let A_1, \dots, A_n be disjoint events that form a partition of the sample space and assume $P(A_i) > 0$ for all i

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B) \\ &= P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n) \\ &= \sum_{i=1}^n P(B|A_i)P(A_i) \\ &= P(B|A)P(A) + P(B|A^C)P(A^C) \end{aligned}$$

- Example: quiz 1-extra

Bayes rule

- Use the conditional probability law and the law of total probability

$$\begin{aligned}P(A|B) &= \frac{P(A \cap B)}{P(B)} \\&= \frac{P(A)P(B|A)}{P(B)} \\&= \frac{P(A)P(B|A)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \\&= \frac{P(A)P(B|A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}\end{aligned}$$

- Example: slide 2-23, quiz 1-3

Independence

- Intuition: Events A and B are independent if knowing whether A occurred provides no information about whether B occurred
- rolling two fair dice, getting one for one die and getting six for the other one

$$P(A \cap B) = P(A)P(B) \Leftrightarrow A \perp\!\!\!\perp B$$

- Implication

$$\begin{aligned}P(A) &= P(A|B) = P(A|B^C) \\P(A \cap B) &= P(A)P(B) = P(A|B)P(B) = P(A|B^C)P(B^C)\end{aligned}$$

- Analogous concept: conditional independence
- Example: slide 1-26, quiz 1A

Random variables and distributions

Random variables

- Definition: A random variable is a real valued function that maps the sample space (Ω) to the real numbers (\mathbb{R})

Probability law

- Definition: The probability law $P_X(x)$ of a random variable X is a real-valued function that assigns probabilities to each of its possible values $x \in X$. i.e.) $P_X : x \rightarrow [0, 1]$. It encodes our knowledge or belief about the likelihood of the outcomes in Ω .
- If random variables are discrete, we call it the probability mass function (PMF)
- If random variables are continuous, we call it the probability density function (PDF)

Cumulative function

- A cumulative distribution function $f_Y(y)$ of a random variable Y is a non-decreasing function that gives you the probability that $Y \leq y$: $F_Y(y) = Pr(Y \leq y)$
- Discrete: $F_Y(y) = \sum_{Y' \leq y} P_Y(Y')$
- Continuous: $F_Y(y) = \int_{-\infty}^y f_Y(Y') dY'$ where $f_Y(y)$ is the PDF for a random variable Y

Expectation and variance

- Summary of probability distribution
- Expectation (average): measure of location

$$\text{Discrete: } \mathbb{E}[X] = \sum_x x P_X(x)$$

$$\text{Continuous: } \mathbb{E}[X] = \int_{-\infty}^x f_X(x) dx$$

- Variance: measure of dispersion

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- Examples: quiz 1-1B

Multiple random variables

Joint PMF and marginal PMF

- $f_{Y|X}(y|x)$: conditional PDF of Y given $X = x$
- $f_Y(y)$: conditional PDF of Y
- Note $\int_x \int_y f_{X,Y}(x,y) dy dx = \int_y \int_x f_{Y|X}(y|x) f_X(x) dx dy = 1$

Conditional expectation

- Discrete: $\mathbb{E}[Y|X = x] = \sum_y y Pr(Y = y|X = x) = \sum_y y Pr_{Y|X}(y|x)$
- Continuous: $\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$
- Note that $\mathbb{E}[Y|X = x]$ is a function of X (i.e. $\mathbb{E}[Y|X = x]$ depends on the value of X)
- Law of total expectation: $\mathbb{E}_Y[Y] = \mathbb{E}_X[\mathbb{E}_Y[Y|X]]$
- Examples: quiz 1-1E

Conditional variance

- Discrete: $\mathbb{V}(Y|X = x) = \sum_y (y - \mathbb{E}[Y|X = x])^2 P_{Y|X}(y|x)$
- Continuous: $\mathbb{V}(Y|X = x) = \int_{-\infty}^{\infty} (y - \mathbb{E}[Y|X = x])^2 P_{Y|X}(y|x) dy$
- Law of total variance: $\mathbb{V}(Y) = \mathbb{E}[\mathbb{V}(Y|X)] + \mathbb{V}(\mathbb{E}[Y|X])$

Independence and covariance

- Independence implies covariance (and hence correlation) equals zero but not the other way around
- Useful pictures on slide 4-31 and proof on 4-28
- Usage of conditional independence: multiple regression

Hypothesis testing (5)

Intuition

- The goal is to estimate the parameters of the population distribution (estimands). However, we only observe a part of random variables that are (supposed to be) generated by the population distribution (sample). We try to come up with functions of sample data (statistics) which we use to make an inference about the estimands (estimators). Estimators will give you particular values that are realized from a given sample (estimates).
- These are our assumptions about how the world works

Properties of estimators

- There are multiple estimators you can choose from.
- Researchers often rely on the distribution of an estimator due to repeated sampling (the sampling distribution)
- Nice summary on slide 5-1-36, examples on the slide 1-2 and 1-4

Finite-sample properties

- Unbiasedness: Is the sampling distribution of our estimator centered at the true parameter value? $\mathbb{E}[\hat{\theta}^*] = \mu$ if and only if bias is zero
- Efficiency: Is the variance of the sampling distribution of our estimator reasonably small? $\mathbb{V}(\hat{\theta}^*) \leq \mathbb{V}(\hat{\theta})$

Asymptotic properties ($n \rightarrow \infty$)

- Consistency: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value? $\hat{\theta}^* \rightarrow \theta$ (Law of large numbers)
- Asymptotic normality: As our sample size grows large, does the sampling distribution of our estimator approach a normal distribution? $\sqrt{n}(\hat{\theta}^* - \theta) \rightarrow \mathcal{N}(0, \sigma^2)$ (Central limit theorem)

Examples of estimators with properties above

- OLS: all of these
- MLE: all but unbiasedness

Mean squared error

- Examine and bias and variance of the estimator

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = (\mathbb{E}[\hat{\theta}] - \theta) + \mathbb{V}(\hat{\theta}) + \text{Bias}(\hat{\theta}) + \text{Variance}(\hat{\theta})$$

Confidence interval

- The sampling distribution of the sample average is $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$
- Note that μ and σ are population parameters and \bar{X}_n is an estimate
- Useful summary on the slide 5-2-22
- Examples: quiz 1-1D, 2-2

Population is normal with know variance (σ^2)

- Derive a 95% confidence interval of a parameter μ

$$\begin{aligned} Pr(\bar{X}_n - 1.96\sigma/\sqrt{n} < \mu < \bar{X}_n + 1.96\sigma/\sqrt{n}) &= .95 \\ \therefore [\bar{X}_n - 1.96\sigma/\sqrt{n}, \bar{X}_n + 1.96\sigma/\sqrt{n}] \end{aligned}$$

- More general form: Let $Z_{\alpha/2}$ be the $100(1/\alpha/2)$ percentile of the standard normal distribution

$$[\bar{X}_n - Z_{\alpha/2}\sigma/\sqrt{n}, \bar{X}_n + Z_{\alpha/2}\sigma/\sqrt{n}]$$

Population is normal with unknow variance

- Use sample variance and the t-distribution
- Sample variance: $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ (unbiased and consistent)
- t-distribution with $n - 1$ degree of freedom: $\tau_{n-1} \sim \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$
- To construct a confidence interval, swap Z with t which comes from t-distribution and σ with S_n

$$[\bar{X}_n - t_{\alpha/2}S_n/\sqrt{n}, \bar{X}_n + t_{\alpha/2}S_n/\sqrt{n}]$$

Unknown distribution

- Use the CLT: $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \rightarrow \mathcal{N}(0, 1)$
- Also use the LLN: $S_n^2 \rightarrow \sigma^2$
- The rest of the procedure is the same as above