Parametric Statistics

Pre-Midterm Recap

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

November 8, 2023

Probability Basics

- Probability spaces formalize uncertainty.
- Sample spaces describe possible outcomes of a random experiment.
- Events are subsets of the sample space.
- Probability measure defines how probable events are.
- We need to define them on sets of events that are *measurable*, i.e., σ-algebras.
- Probability measures need to satisfy the *axioms of probability.*
- In discrete uniform sample spaces, we can compute probabilities by counting (a) the number of outcomes in our event (b) the number of outcomes in the sample space.

Independence/Conditional Probability/Bayes Rule

- Two events are called independent when knowing the value of one doesn't influence the probability of the value of the other.
- The conditional probability of A given B denotes the probability of event A in a world where B has occurred.
- We can use the multiplication to compute conditional, marginal and joint probabilities.
- Bayes rule connects $P(A|B)$ and $P(B|A)$. These two are confused but they are not the same.

Joint Distributions

- Two random variables have a bivariate joint distribution.
- More than two RVs have a multivariate joint distribution.
- We can compute marginal, conditional distributions from the joint pf.
- Independence is defined for RVs.
- Functions of RVs are RVs.

Conditional Probability

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table displays the distribution of health status of respondents to this survey (excellent, very good, good, fair, poor) and whether or not they have health insurance.

		Health Status				
		Excellent	Very good	Good	Fair	Poor
Health	No	0.0230	0.0364	0.0427	0.0192	0.0050
Coverage	Yes	0.2099	0.3123	0.2410	0.0817	0.0289

1. Are being in excellent health and having health coverage mutually exclusive?
2. What is the probability that a randomly chosen individual has excellent health?
3. What is the probability that a randomly chosen individual has excellent health given that he has health coverage?
4. What is the probability that a randomly chosen individual has excellent health given that he doesn't have health coverage?
5. Do having excellent health and having health coverage appear to be independent?

Solution

1. No, there are individuals who are both excellent in health and have health coverage.
2. P(excellent health) = 0.2329
3. P(excellent health | health coverage) = 0.2099/0.8738 = 0.24
4. P(excellent health | no health coverage) = 0.0230/0.1262 = 0.18
5. No, because the probability that a person has excellent health varies between the two health coverage categories (24% vs 18%). That is, knowing something about someone's health coverage provides useful information in predicting whether the person has excellent health, which means the variables are not independent.

Example

- Defaulting on a loan means failing to pay back on time. The default rate among a bank's clients is 1%. You develop a test to predict which clients will default and which will not. Your test gives 4% false positives, i.e., predicts a client will default who in fact will not. The test makes no false negatives, i.e., it never predicts that a client will not default when in fact they will.
    1. Suppose a client tests positive. What is the probability that they will default?
    2. Someone offers to bet you that the client you tested above will not default. They want you to pay them 100 euro if the client does not default, and they will pay you 400 euro if the client defaults. Should you take the bet?

Expectations

- Expectation is a summary of a distribution.
- We can compute the expectation of a function of an RV using LOTUS.
- Expectations are linear.
- Variance is a summary of how spread out a distribution is.
- Covariance describes how much two variables vary together.
- Correlation is covariance without scale.

Conditional Expectation/Moments

- Conditional Expectation, Conditional Variance are functions of the conditioning variable.
- Law of total expectation/law of total variance can help us compute variances and expectations of complex functions.
- The means of powers X^k of an RV X are called moments of X.
- They can help us derive distributions of sums of independent random variables and prove limiting properties of distributions.

Special Distributions

Distribution	PMF/PDF and Support	Expected Value	Variance	MGF
Bernoulli Bern(p)	$P(X = 1) = p$ $P(X = 0) = q = 1 - p$	p	pq	$q + pe^t$
Binomial Bin(n, p)	$P(X = k) = \binom{n}{k} p^k q^{n-k}$ $k \in \{0, 1, 2, \ldots n\}$	np	npq	$(q + pe^t)^n$
Geometric Geom(p)	$P(X = k) = q^k p$ $k \in \{0, 1, 2, \ldots\}$	q/p	q/p^2	$\frac{p}{1-qe^t}$, $qe^t < 1$
Poisson Pois(λ)	$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ $k \in \{0, 1, 2, \ldots\}$	λ	λ	$e^{\lambda(e^t - 1)}$
Uniform Unif(a, b)	$f(x) = \frac{1}{b-a}$ $x \in (a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
Normal (μ, σ^2)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2(2\sigma^2)}$ $x \in (-\infty, \infty)$	μ	σ^2	$e^{t\mu + \frac{\sigma^2 t^2}{2}}$
Exponential Expo(λ)	$f(x) = \lambda e^{-\lambda x}$ $x \in (0, \infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$, $t < \lambda$

Large Random Samples

- Inequalities (Markov, Chebysev) bound how far away a random variable can be from its mean.
- The average of a random sample of i.i.d. random variables is called their sample mean.
- WLLN: The sample mean converges in probability to the true distribution mean.
- CLT: The sample mean is asymptotically normal.

Example

- You are doing a poll on "ratio of people who believe in climate change".
- True ratio: p, estimate $\overline{X}_n$.
- Your boss wants a guarantee that your prediction will be "off" by more than 1% with probability at most 5%

Example

- You are doing a poll on "ratio of people who believe in climate change".
- True ratio: p, estimate $\overline{X}_n$.
- Your boss wants a guarantee that your prediction will be "off" by more than 1% with probability at most 5%
- No guarantee for finding exactly p, so

$$P(|\overline{X}_n - p| \geq 0.01) \leq 0.05$$

- Apply Chebysev inequality with $t = 0.01$:
- Apply CLT:

Bayesian Inference

- Pick a prior distribution.
- Compute the likelihood.
- Use Bayes' theorem to compute the posterior distribution:

$$\text{Posterior Distribution} \propto \text{Likelihood} \times \text{Prior Distribution}$$

- Perform Sensitivity Analysis.
- Summarize the posterior distribution:
  - Posterior mean minimizes squared error loss.
  - Posterior median minimizes absolute error loss.

MLE estimation

- Find the likelihood function.
- Find the log likelihood function.
- Take the derivative to find the global optimum $\hat{\theta}$
- Use the second derivative to check that $\hat{\theta}$ is a maximizer.

Example

The data

$$\mathbf{x} = \{1, 16, 13, 9, 30, 6, 2, 21, 1\}$$

are an i.i.d. sample from a distribution with p.m.f:

$$\mathrm{P}_\theta(X = x) = (1-\theta)\theta^x, x \in = \{0, 1, 2, \ldots\}$$

for some $\theta \in (0, 1)$ [1]

1. Find the likelihood function and MLE for θ.
2. Find the method of moments estimator for θ.
3. Using a $Beta(\alpha, \beta)$ prior for θ, find the posterior distribution $f(\theta|\mathbf{x})$.

Useful formulae

Beta distribution: $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$ where $\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and Γ is the Gamma function.
Geometric distribution: $f(x) = (1-p)^x p$, $E[x] = \frac{1-p}{p}$

[1] this is the Geometric distribution but parameterized in terms of the failure probability instead of the more usual success probability.