Statistical Theory II
Practical Labs

Lab 2: Recap — Bayesian Inference for Discrete Random Variables

Paulius Kazlauskas

Outline

Discrete Random Variables

Bayesian Inference

Named Distributions

R Reference

Discrete Random Variables

Probability Mass Function

Definition

A function $f(y_i) = P(Y = y_i)$ is a valid PMF if:

$$f(y_i) \geq 0 \qquad \text{and} \qquad \sum_i f(y_i) = 1$$

Computing Probabilities Over a Range

$$P(a < Y < b) \; = \; \sum_{\{i \,:\, a < y_i < b\}} f(y_i)$$

Today's example

$P(1 < Y < 5) = P(Y = 2) + P(Y = 3) + P(Y = 4) = 0.2 + 0.1 + 0.2 = 0.5$

Expected Value

Definitions

$$E(Y) = \sum_i y_i \, f(y_i) \qquad E(Y^2) = \sum_i y_i^2 \, f(y_i)$$

Linearity

$$E(aY + b) = a\,E(Y) + b$$

Additivity (always)

$$E(Y + W) = E(Y) + E(W)$$

Note

Additivity holds **regardless** of whether Y and W are independent.

Variance

Shortcut Formula

$$\mathrm{Var}(Y) = E(Y^2) - \left[E(Y)\right]^2$$

Scaling Rule (always)

$$\mathrm{Var}(aY + b) = a^2\,\mathrm{Var}(Y)$$

Constants vanish; shifts don't matter.

Sum Rule (independent only)

$$\mathrm{Var}(Y + W) = \mathrm{Var}(Y) + \mathrm{Var}(W)$$

Dependent case — substitute first

If $W = 2Y + 5$: $\mathrm{Var}(Y + W) = \mathrm{Var}(3Y + 5) = 9\,\mathrm{Var}(Y)$

Rules at a Glance

Result	Formula	Always?
$E(aY + b)$	$a\,E(Y) + b$	$\checkmark$
$E(Y + W)$	$E(Y) + E(W)$	$\checkmark$
$\mathrm{Var}(aY + b)$	$a^2\,\mathrm{Var}(Y)$	$\checkmark$
$\mathrm{Var}(Y + W)$	$\mathrm{Var}(Y) + \mathrm{Var}(W)$	$Y \perp W$ only
$\mathrm{Var}(Y + W)$	Substitute W, then apply scaling rule	$\checkmark$

Bayesian Inference

The Core Formula

$$\underbrace{P(\theta \mid y)}_{\text{Posterior}} \propto \underbrace{P(y \mid \theta)}_{\text{Likelihood}} \times \underbrace{P(\theta)}_{\text{Prior}}$$

Normalised Posterior

$$P(\theta_i \mid y) = \frac{P(y \mid \theta_i)\ P(\theta_i)}{\sum_j P(y \mid \theta_j)\ P(\theta_j)}$$

The denominator is the *marginal likelihood* — it ensures the posterior sums to 1.

Five Steps to a Posterior

1. **List** all possible values of the parameter θ.
2. **Assign priors** $P(\theta_i)$ — your belief *before* seeing any data.
3. **Compute likelihoods** $P(y \mid \theta_i)$ — how probable is the observed y under each θ_i?
4. **Multiply** prior $\times$ likelihood for each θ_i.
5. **Normalise** by dividing each product by the column total.

Practical tip

Build a table with columns: *prior*, *likelihood*, *prior* $\times$ *likelihood*, *posterior*. One row per value of θ.

Two Invariance Properties

Scaling the Prior

Replace $P(\theta_i)$ with $c \cdot P(\theta_i)$. The c cancels in numerator and denominator:

$$\frac{c\, P(\theta_i)\, P(y \mid \theta_i)}{\sum_j c\, P(\theta_j)\, P(y \mid \theta_j)} = \frac{P(\theta_i)\, P(y \mid \theta_i)}{\sum_j P(\theta_j)\, P(y \mid \theta_j)}$$

Scaling the Likelihood

Replace $P(y \mid \theta_i)$ with $c \cdot P(y \mid \theta_i)$. Identical cancellation applies.

Why it matters

Priors and likelihoods only need to be **proportional** to valid probabilities.

Unnormalised weights are fine as inputs.

Sequential vs. Single-Step Updating

Sequential

1. Compute $P(\theta \mid y_1)$.
2. Use it as the new prior.
3. Update on y_2 to get $P(\theta \mid y_1, y_2)$.

Single Step

Use the original prior and the **joint** likelihood:

$$P(\theta \mid y_1, y_2) \propto P(y_1, y_2 \mid \theta)\, P(\theta)$$

Key result

Both approaches give **identical** posteriors. This follows from the chain rule of probability.

Without vs. With Replacement: Likelihoods

Box with 4 balls, x of which are red. Draw 2 red balls.

Without Replacement

Draws are **dependent**:

$$P(R_1, R_2 \mid x) = \frac{x}{4} \cdot \frac{x-1}{3}$$

One fewer red ball and one fewer total after draw 1.

With Replacement

Draws are **independent**:

$$P(R_1, R_2 \mid x) = \left(\frac{x}{4}\right)^2$$

Box is reset before draw 2.

Rule of thumb

Ask: does the first draw change what is left in the box for the second?

Named Distributions

Binomial Distribution

Setup

$Y \sim \mathrm{Binomial}(n, \pi)$ — number of successes in n independent trials, success probability π.

PMF

$$P(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$$

R Functions

`dbinom(y,n,p)`	$P(Y = y)$
`pbinom(y,n,p)`	$P(Y \leq y)$
`qbinom(p,n,p)`	quantile
`rbinom(s,n,p)`	simulate

Poisson Distribution

Setup

$Y \sim \mathrm{Poisson}(\mu)$ — count of events in a fixed interval, mean rate $\mu > 0$.

PMF

$$P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}$$

R Functions

`dpois(y,mu)`	$P(Y = y)$
`ppois(y,mu)`	$P(Y \leq y)$
`qpois(p,mu)`	quantile
`rpois(s,mu)`	simulate

R Reference

R: Discrete Distribution Calculations

```r
y <- c(1, 2, 3, 4, 5)
py <- c(0.1, 0.2, 0.1, 0.2, 0.4)

sum(py[y > 1 & y < 5]) # P(1 < Y < 5)

my <- sum(y * py) # E(Y)
vary <- sum(y^2 * py) - my^2 # Var(Y) -- subtract E(Y)^2

# Simulate and verify
ys <- sample(y, size = 1e5, replace = TRUE, prob = py)
mean(ys) # approx E(Y)
var(ys) # approx Var(Y)
```

R: Generic Bayesian Updating

```
# Works for any model -- just supply prior and likelihood vectors
bayes_update <- function(prior, likelihood) {
  w <- prior * likelihood
  return(w / sum(w))
}

# Poisson example: y = 1 observed
mu <- c(1, 1.5, 2, 2.5)
prior <- c(2/6, 1/6, 2/6, 1/6)
posterior <- bayes_update(prior, dpois(1, mu))

# Binomial example: 2 heads in 3 flips
pi_vals <- c(0.3, 0.5)
prior_pi <- c(0.6, 0.4)
posterior <- bayes_update(prior_pi, dbinom(2, 3, pi_vals))
```

Summary

Expectation & Variance

$E(aY + b) = aE(Y) + b$

$E(Y + W) = E(Y) + E(W)$ *always*

$\mathrm{Var}(aY + b) = a^2\mathrm{Var}(Y)$

$\mathrm{Var}(Y + W)$: additive only if $Y \perp W$

Distributions

`dbinom(y, n, pi)`

`dpois(y, mu)`

Bayesian Inference

Posterior $\propto$ Prior $\times$ Likelihood

Scaling prior or likelihood: no effect

Sequential $=$ Single-step updating

Replacement changes the likelihood

In R

`sample()` to simulate

`mean()`, `var()` on simulated draws

`w / sum(w)` to normalise