Math3810 - Probability
Section 001 - Fall 2025
Notes: Bayesian Updating

University of Colorado Denver / College of Liberal Arts and Sciences

Department of Mathematics - Dr. Robert Rostermundt

The Problem:

We have an unknown parameter θ satisfying $0 \leq \theta \leq 2\pi$ and we want to use empirical data to make probabilistic arguments about the true value of θ. So let r be our observation where there is some random noise $n \sim N(0, \sigma^2)$. So $\theta = r + n$. Assuming we have no prior knowledge about θ we will start with the prior distribution $\theta \sim Uni[0, 2\pi]$. We want to determine the conditional distribution $f(\theta \,|\, r)$. This is our posterior distribution reflecting our updated beliefs about θ given the observed data.

Theoretical Tools:

Here are the basic tools for this problem.

- **Continuous Bayes' Rule**: Let X, Y be continuous jointly distributed random variables. Then

$$f_{Y|X}(y, x) = \frac{f_Y(y) \cdot f_{X|Y}(x, y)}{\displaystyle\int_{-\infty}^{\infty} f_Y(y) \cdot f_{X|Y}(x, y) \, dy}$$

  We are not providing a derivation of this rule, but this can be found in most textbooks on the subject. Noticing that the denominator is constant we will often write

$$f_{Y|X}(y, x) \propto f_Y(y) \cdot f_{X|Y}(x, y)$$

  where the symbol $\propto$ means "proportional to."

- **Transformation of** $N(\mu, \sigma^2)$: Let $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$. Then $Y \sim N(a\mu + b, a^2\sigma^2)$. This is a result that we proved in class. This means, in our example, that we can easily describe the following conditional distribution:

$$r \,|\, \theta \sim N(\theta, \sigma^2).$$

  To see this, $r = -n + \theta$ is a transformation of $an + b$ the normal random variable $n \sim N(0, \sigma^2)$ with $a = -1$ and $b = \theta$. So we have

$$f_{r|\theta}(r, \theta) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(r - \theta)^2}{2\sigma^2}\right]$$

Calculations:

We start with the formula for continuous Bayes' Rule.

$$f_{\theta|r}(\theta, r) = \frac{f_\theta(\theta) \cdot f_{r|\theta}(r, \theta)}{\displaystyle\int_{-\infty}^{\infty} f_\theta(\theta) \cdot f_{r|\theta}(r, \theta)\, d\theta} = \frac{\dfrac{1}{2\pi} \cdot \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(r-\theta)^2/2\sigma^2}}{\dfrac{1}{\sigma(2\pi)^{3/2}} \displaystyle\int_{\theta=0}^{\theta=2\pi} e^{-(r-\theta)^2/2\sigma^2}\, d\theta}$$

So we have

$$f_{\theta|r}(\theta, r) \propto e^{-(r-\theta)^2/2\sigma^2}$$

where $0 \le \theta \le 2\pi$. This is a truncated, scaled, density of a normal distribution $N(\theta, \sigma^2)$. We will lot this in the simulation section below.

We can continue to update by using the above posterior distribution as our new prior distribution. For various reasons, we would typically denote $f_{r|\theta}(r, \theta) = L(r \mid \theta)$ the "likelihood function." We now simplify notation and let

$$f_1(\theta \mid r_1) = \frac{f(\theta) \cdot L(r \mid \theta)}{\displaystyle\int_{-\infty}^{\infty} f(\theta) \cdot L(r \mid \theta)\, d\theta} = \frac{\dfrac{1}{2\pi} \cdot \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(r_1-\theta)^2/2\sigma^2}}{\dfrac{1}{\sigma(2\pi)^{3/2}} \displaystyle\int_{\theta=0}^{\theta=2\pi} e^{-(r_1-\theta)^2/2\sigma^2}\, d\theta}$$

Next we make observation r_2 to get

$$f_2(\theta \mid r_1, r_2) \propto f_1(\theta \mid r_1) L(r_2 \mid \theta).$$

That is,

$$f_2(\theta \mid r_1, r_2) = \frac{f_1(\theta \mid r_1) L(r_2 \mid \theta)}{\displaystyle\int_{\theta=0}^{\theta=2\pi} f(\theta \mid r_1) L(r_2 \mid \theta)\, d\theta}$$

$$f_2(\theta \mid r_1, r_2) \propto exp\left[-\frac{(r_1 - \theta)^2}{2\sigma^2}\right] \times exp\left[-\frac{(r_2 - \theta)^2}{2\sigma^2}\right] = exp\left[-\frac{(r_1 - \theta)^2 + (r_2 - \theta)^2}{2\sigma^2}\right].$$

After completing the square we get

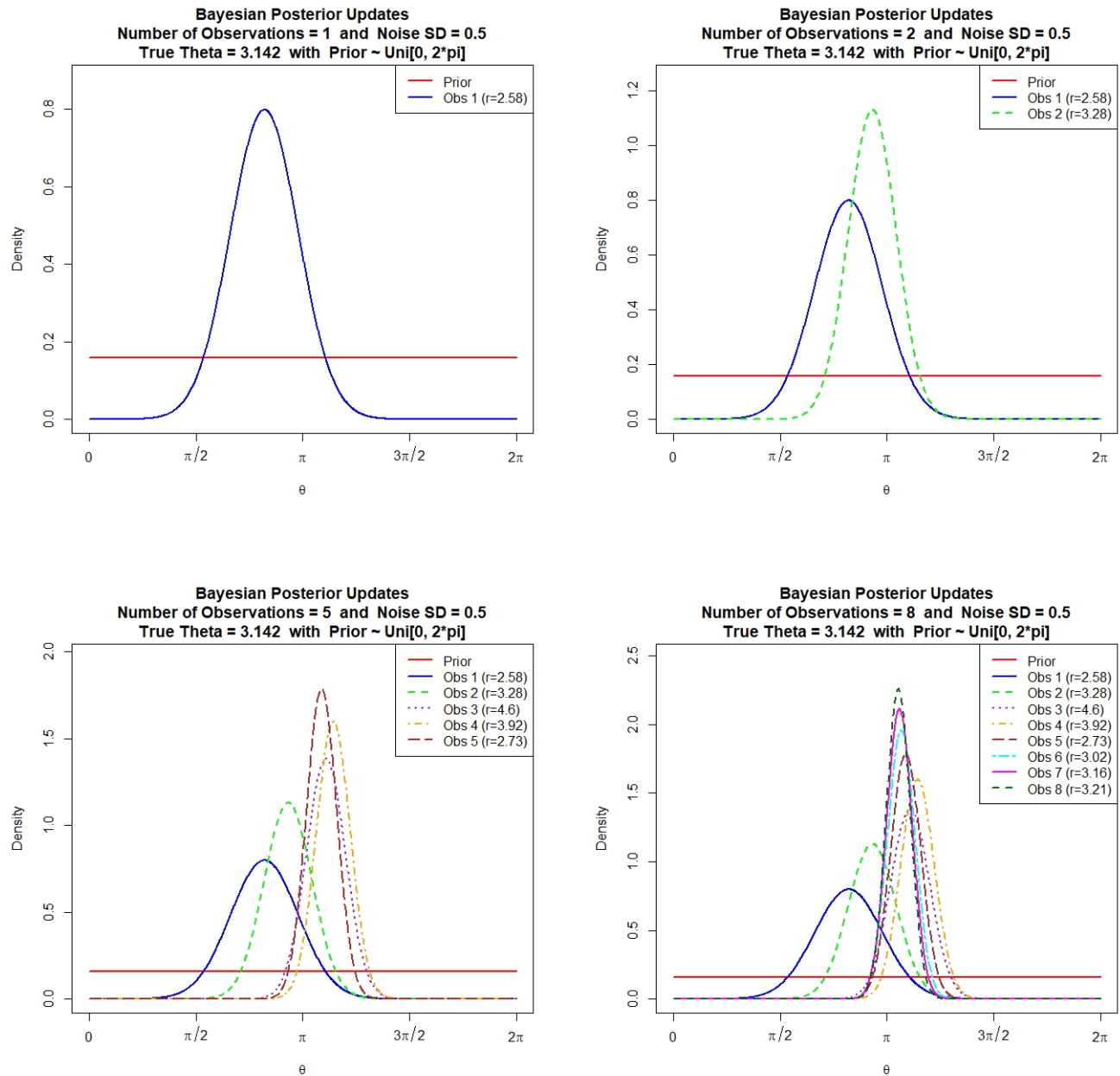$$f_2(\theta \mid r_1, r_2) \propto exp\left[-\frac{(\theta - \bar{r}_2)^2}{2\sigma^2}\right],$$

where $\bar{r}_2 = (r_1 + r_2)/2$ is the average of the previous two observations r_1 and r_2. Hence $f_2(\theta \mid r_1, r_2)$ is a truncated normal centered at the average of the previous observations, with variance $\sigma^2/2$. If we continue we will have

$$\theta \mid r_1, r_2, \cdots, r_n \sim N\left(\bar{r}_n, \frac{\sigma^2}{n}\right).$$

centered at the average of all observations and truncated to $[0, 2\pi]$.

Simulations:

I have simulated this process in R. Below are graphics from the same overall simulation showing the number updates $n = 1, 2, 5, 8$. The R code is provided the end of the document.

One thing to notice is that the variance of the normal posterior distribution (before truncation and normalization) is σ^2/n. We can clearly see this with each updated posterior.

The R Code:

Here is the R code used for the above simulations.

```r
# Simulated Bayesian Updating

# ----------------------------
# User settings
# ----------------------------

# Option 1: manually specify observed values
#r_obs <- c(2.58, 3.28, 4.6, 3.92, 2.73)
#r_obs <- c(2.58, 3.28, 4.6, 3.92, 2.73, 3.02, 3.16, 3.21)


# Option 2: simulate n_trials from a true theta
theta_true <- pi          # true underlying theta (radians)
sigma <- 0.5              # noise standard deviation
n_trials <- 8            # number of trials
r_obs <- theta_true + rnorm(n_trials, mean=0, sd=sigma) # generate noisy observations


# ----------------------------
# Setup
# ----------------------------
theta_grid <- seq(0, 2*pi, length.out = 1000)
phi <- function(x) dnorm(x)
prior_vals <- rep(1/(2*pi), length(theta_grid))   # uniform prior


# Safe, visually distinct colors for posteriors
posterior_colors <- c("blue", "green", "purple", "orange", "brown", "cyan",
"magenta", "darkgreen", "darkred", "darkblue")
line_types <- 1:length(r_obs)

# ----------------------------
# Compute sequential posteriors and store
# ----------------------------
posterior_vals <- prior_vals
all_posteriors <- list(prior_vals)  # include prior

for (r in r_obs) {
  posterior_vals <- posterior_vals*phi((r - theta_grid)/sigma)
  posterior_vals <- posterior_vals/sum(posterior_vals)*length(theta_grid)/(2*pi)
  all_posteriors <- append(all_posteriors, list(posterior_vals))
}

# ----------------------------
# Determine max density for plotting
# ----------------------------
ymax <- max(sapply(all_posteriors, max)) * 1.1

# ----------------------------
```

```r
# Detailed plot title
# ----------------------------
plot_title <- paste0(
  "Bayesian Posterior Updates\n",
  "Number of Observations = ", length(r_obs),
  "  and  Noise SD = ", sigma, "\n",
  "True Theta = ", round(theta_true, 3),
  "  with  Prior ~ Uni[0, 2*pi]"
)

# ----------------------------
# Base R plot
# ----------------------------
plot(theta_grid, prior_vals, type='l', col='red', lwd=2,
     ylim=c(0, ymax),
     xaxt='n', xlab=expression(theta), ylab='Density',
     main=plot_title)

# x-axis multiples of pi
axis(side=1,
     at=c(0, pi/2, pi, 3*pi/2, 2*pi),
     labels=c(expression(0), expression(pi/2), expression(pi),
              expression(3*pi/2), expression(2*pi)))

# Plot sequential posteriors
for (i in seq_along(r_obs)) {
  lines(theta_grid, all_posteriors[[i + 1]], col=posterior_colors[i],
  lwd=2, lty=line_types[i])
}

# ----------------------------
# Add legend with observation values
# ----------------------------
legend_labels <- c("Prior", paste0("Obs ", seq_along(r_obs),
" (r=", round(r_obs, 2), ")"))
legend("topright",
       legend=legend_labels,
       col=c("red", posterior_colors[1:length(r_obs)]),
       lwd=2, lty=c(1, line_types))
```

Please let me know if you have any questions, comments, or corrections!