Solution 1

(a) We assume throughout:

$$Y \mid \mu \sim \mathcal{N}(\mu, 1), \quad \mu \sim \mathcal{N}(0, 1), \quad \tau \in \{-1, 1\} \text{ independent of } (\mu, Y).$$

Then marginally, $Y \sim \mathcal{N}(0, 2)$. as shown during the lecture.

- For the unfocused forecast F_1, we have that the predictive distribution is:

$$F_1 = \tfrac{1}{2}\mathcal{N}(\mu, 1) + \tfrac{1}{2}\mathcal{N}(\mu + \tau, 1).$$

Marginal calibration We compute the expected forecast cdf:

$$\begin{aligned} \mathrm{E}[F_1(y)] &= \mathbb{E}_{\mu,\tau}\left[\tfrac{1}{2}\Phi\left(\tfrac{y-\mu}{1}\right) + \tfrac{1}{2}\Phi\left(\tfrac{y-\mu-\tau}{1}\right)\right] \\ &= \tfrac{1}{2}\mathbb{E}_{\mu}\left[\Phi(y-\mu)\right] + \tfrac{1}{4}\mathbb{E}_{\mu}\left[\Phi(y-\mu-1) + \Phi(y-\mu+1)\right], \end{aligned}$$

where the last equality is obtained using the fact that $\tau \in \{-1, 1\}$ with equal probability and independently from μ. Thus,

$$\mathrm{E}_{\mu,\tau}\left\{\Phi(y-\mu-\tau)\right\} = \tfrac{1}{2}\mathrm{E}_{\mu}\left\{\Phi(y-\mu-1)\right\} + \tfrac{1}{2}\mathrm{E}_{\mu}\left\{\Phi(y-\mu+1)\right\},$$

Using derivations similar to the ones we used during the lecture to derive the marginal distribution of Y, i.e., derivations of a convolution of two normals, we can show that $\mathrm{E}_{\mu}\{\Phi(y - \mu - cst)\} = \Phi_{cst,2}(y)$, i.e. a Gaussian with mean the constant cst and variance 2 (that of Y). Thus,

$$\mathrm{E}[F_1(y)] = \tfrac{1}{2}\Phi_{0,2}(y) + \tfrac{1}{4}\left\{\Phi_{1,2}(y) + \Phi_{-1,2}(y)\right\}$$

The expression of $\mathrm{E}[F_1(y)]$ differs from $\mathbb{P}(Y \leq y) = \Phi_{0,2}(y)$.

Probabilistic calibration Define the probability integral transform:

$$Z = F_1(Y) = \tfrac{1}{2}\Phi(Y-\mu) + \tfrac{1}{2}\Phi(Y-\mu-\tau).$$

Since $Y \mid \mu \sim \mathcal{N}(\mu, 1)$, $Y - \mu \sim \mathcal{N}(0, 1)$. Hence:

$$Z = \tfrac{1}{2}\Phi(Z_1) + \tfrac{1}{2}\Phi(Z_1 - \tau), \quad \text{where } Z_1 \sim \mathcal{N}(0, 1), \tau \in \{-1, 1\}.$$

Due to the added noise from the random shift τ, the PIT is smoothed, and simulations show that $Z \sim \text{Uniform}(0, 1)$ approximately. Thus, the forecast F_1 is probabilistically calibrated.

- For the sign-reversed forecast F_2, the predictive distribution is:

$$F_2 = \mathcal{N}(-\mu, 1).$$

Marginal calibration

$$\begin{aligned} \mathbb{E}[F_2(y)] = \mathbb{E}_{\mu}\left[\Phi\left(\tfrac{y+\mu}{1}\right)\right] = \mathbb{P}(Z \leq y), &\quad \text{where } Z = -\mu + \varepsilon, \varepsilon \sim \mathcal{N}(0, 1) \\ \Rightarrow Z \sim \mathcal{N}(0, 2) \Rightarrow \mathbb{E}[F_2(y)] = \Phi_{0,2}(y) = \mathbb{P}(Y \leq y).& \end{aligned}$$

Thus, the forecast F_2 is marginally calibrated.

Probabilistic Calibration. The PIT is:

$$Z = F_2(Y) = \Phi(Y + \mu).$$

Since $Y \sim \mathcal{N}(\mu, 1) \Rightarrow Y + \mu \sim \mathcal{N}(2\mu, 1)$, the distribution of Z depends on μ and hence is not uniform. Thus, the forecast F_2 is not probabilistically calibrated.

(b)
```
set.seed(22)
N <- 10000
mu <- rnorm(N)
tau <- sample(c(-1, 1), N, replace = TRUE)
Y <- rnorm(N, mean = mu, sd = 1)

# compute PITs
# F1 <- pnorm(Y, mean = mu, sd = 1)
# F2 <- pnorm(Y, mean = 0, sd = sqrt(2))
F1 <- 0.5 * pnorm(Y, mean = mu, sd = 1) +
      0.5 * pnorm(Y, mean = mu + tau, sd = 1)
F2 <- pnorm(Y, mean = -mu, sd = 1)

par(mfrow = c(1, 2))
# hist(F1, breaks = 20, main = "PIT: Perfect", xlab = "PIT", freq = FALSE)
# hist(F2, breaks = 20, main = "PIT: Climatological", xlab = "PIT", freq = FALSE)
hist(F1, breaks = 20, main = "PIT: Unfocused", xlab = "PIT", freq = FALSE)
hist(F2, breaks = 20, main = "PIT: Sign-Reversed", xlab = "PIT", freq = FALSE)

### Assess marginal calibration empirically
y_grid <- seq(-5, 5, length.out = 200)

# Empirical CDF of Y
ecdf_Y <- ecdf(Y)
empirical_probs <- sapply(y_grid, ecdf_Y)

# avgF1 <- sapply(y_grid, function(y) mean(pnorm(y, mean = mu, sd = 1)))
# avgF2 <- sapply(y_grid, function(y) mean(pnorm(y, mean = 0, sd = sqrt(2))))
avgF1 <- sapply(y_grid, function(y) mean(0.5 * pnorm(y, mean = mu, sd = 1) +
                                          0.5 * pnorm(y, mean = mu + tau, sd = 1)))
avgF2 <- sapply(y_grid, function(y) mean(pnorm(y, mean = -mu, sd = 1)))

plot(y_grid, empirical_probs, type = "l", lwd = 2, col = "black", ylim = c(0,1),
     ylab = "CDF", xlab = "y", main = " ")
# lines(y_grid, avgF1, col = "blue", lty = 2)
# lines(y_grid, avgF2, col = "green", lty = 3)
lines(y_grid, avgF1, col = "red", lty = 4)
lines(y_grid, avgF2, col = "purple", lty = 5)
# legend("bottomright", legend = c("Empirical CDF", "F1 (Perfect)", "F2 (Climatological)",
#                                   "F3 (Unfocused)", "F4 (Sign-Reversed)"),
#        col = c("black", "blue", "green", "red", "purple"), lty = 1:5, cex = 0.8)
legend("bottomright", legend = c("Empirical CDF",
                                 "F1 (Unfocused)", "F2 (Sign-Reversed)"),
       col = c("black", "red", "purple"), lty = 1:5, cex = 0.8)
```

(c) Sharpness refers to the concentration (narrowness) of predictive distributions, independent of calibration and of the distribution of the observation. We will compare the sharpness of both forecasts by comparing their unconditional variances.

- The forecast distribution F_1 is a mixture of two normal distributions with equal variance but different means. Specifically, it is defined as

  $$F_1 = \tfrac{1}{2}\mathcal{N}(\mu, 1) + \tfrac{1}{2}\mathcal{N}(\mu + \tau, 1),$$

  where $\mu \sim \mathcal{N}(0, 1)$ and $\tau \in \{-1, +1\}$ is independent of μ with equal probability.
  Given μ and τ, the conditional mean of $Y' \sim F_1$ is

  $$\mathrm{E}(Y' \mid \mu, \tau) = \tfrac{1}{2}\mu + \tfrac{1}{2}(\mu + \tau) = \mu + \tfrac{\tau}{2},$$

  and the conditional variance is

  $$\mathrm{var}(Y' \mid \mu, \tau) = 1 + \tfrac{1}{4} = \tfrac{5}{4},$$

  using the formula for the variance of a mixture of two Gaussians with equal variances $\sigma^2 = 1$ and means μ and $\mu + \tau$.
  We then apply the law of total variance:

  $$\mathrm{var}(Y') = \mathrm{E}[\mathrm{var}(Y' \mid \mu, \tau)] + \mathrm{var}(\mathrm{E}[Y' \mid \mu, \tau]).$$

  The first term is simply $\mathrm{E}[\mathrm{var}(Y' \mid \mu, \tau)] = \frac{5}{4}$. For the second term, since $\mathrm{E}[Y' \mid \mu, \tau] = \mu + \frac{\tau}{2}$, and μ and τ are independent with the same variance of 1, we get:

  $$\mathrm{var}(\mu + \tau/2) = \mathrm{var}(\mu) + \mathrm{var}(\tau/2) = 1 + \tfrac{1}{4} = \tfrac{5}{4}.$$

  Therefore, the total variance of F_1 is

  $$\mathrm{var}(Y') = \tfrac{5}{4} + \tfrac{5}{4} = \tfrac{5}{2} = 2.5.$$

- The forecast F_2 is

  $$F_2 = \mathcal{N}(-\mu, 1),$$

  where $\mu \sim \mathcal{N}(0, 1)$. Thus, the unconditional distribution of $Y'' \sim F_2$ is obtained by integrating out μ and corresponds to $\mathcal{N}(0, 2)$ (same derivations as the ones we did in the lecture). Thus, $\mathrm{var}(Y'') = 2$.

Hence, the forecast F_2 is sharper than F_1, even though it may be miscalibrated (e.g., predicting the wrong sign).

Solution 2 We will show that the linear score is *not* proper by constructing a counterexample where a forecast $p \neq q$ yields a better (i.e., lower) expected score under the true distribution q.

Let the true density q be the standard normal:

$$q(y) = \tfrac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

Let the forecast p be the uniform distribution on the interval $(-\epsilon, \epsilon)$, where $\epsilon > 0$. Then:

$$p(y) = \begin{cases} \frac{1}{2\epsilon} & \text{if } y \in (-\epsilon, \epsilon), \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{aligned} \mathrm{LinS}(p, q) - \mathrm{LinS}(q, q) &= \int q(y)^2 dy - \tfrac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} q(y) dy \\ &= \tfrac{1}{2\pi} \int e^{-y^2} dy - \tfrac{1}{\sqrt{2\pi}} \tfrac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} e^{-y^2/2} dy \\ &= \tfrac{1}{\sqrt{2\pi}} \left(1/\sqrt{2} - \tfrac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} e^{-y^2/2} dy \right) \end{aligned}$$

The last quantity will be negative for small enough $\epsilon > 0:$ in particular, for $\epsilon < \sqrt{\log 2}$, the integrand in the above satisfies $e^{-y^2/2} > 1/\sqrt{2}$, so

$$\mathrm{LinS}(p, q) - \mathrm{LinS}(q, q) < \tfrac{1}{\sqrt{2\pi}} \left(1/\sqrt{2} - \tfrac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} 1/\sqrt{2} dy\right) < 0$$

This shows that the forecast p performs better than the true distribution q under the linear score. Thus, the linear score is *not proper*, because a forecaster can achieve a better expected score by forecasting a distribution $p \neq q$. It therefore does not incentivise truthful reporting of the forecasters belief and is not used in practice.

Solution 3

(a)
- First note that

$$\begin{aligned} |X - y| &= \int_{-\infty}^{+\infty} I(X \leq u \leq y) du + \int_{-\infty}^{+\infty} I(y \leq u \leq X) du \\ &= \int_{-\infty}^{y} I(X \leq u) du + \int_{y}^{+\infty} I(u \leq X) du. \end{aligned}$$

  Then, applying Fubini to switch expectation and integration, we get

$$\begin{aligned} \mathrm{E}|X - y| &= \int_{-\infty}^{y} \mathrm{E}\{I(X \leq u)\} du + \int_{y}^{\infty} \mathrm{E}\{I(u \leq X)\} du \\ &= \int_{-\infty}^{y} F(u) du + \int_{y}^{+\infty} \{1 - F(u)\} du. \end{aligned}$$

- We use the same trick, i.e.,

$$|X - X'| = \int_{-\infty}^{+\infty} [I(X \leq u \leq X') + I(X' \leq u < X)] du$$

  and the fact that X and X' are independent and identically distributed.
- Expanding the initial expression of the CRPS, we get

$$\begin{aligned} \mathrm{CRPS}(F, y) &= \int_{-\infty}^{\infty} \left(F(z)^2 - 2F(z)\mathbb{I}\{y \leq z\} + \mathbb{I}\{y \leq z\}\right) dz \\ &= \int_{-\infty}^{\infty} F(z)^2 dz - 2 \int_{y}^{\infty} F(z) dz + \int_{y}^{\infty} 1 \, dz. \end{aligned}$$

  This is exactly what we get by computing $\mathrm{E}_F(|X - y|) - \frac{1}{2}\mathrm{E}_F(|X - X'|)$ using the two results derived above.

(b) Let:

$$F(x) = \sum_{m=1}^{N} \omega_m \Phi\left(\tfrac{x - \mu_m}{\sigma_m}\right),$$

where $\omega_m > 0$, $\sum \omega_m = 1$, $\mu_m \in \mathbb{R}$, $\sigma_m > 0$.

Let $A(\mu, \sigma; a) = \mathbb{E}|X - a|$, for $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, using the expression of the expectation of the positive part of a random variable we derived above, we can show that $A(\mu, \sigma; a)$ satisfies:

$$A(\mu, \sigma; a) = 2\sigma\varphi\left(\tfrac{a - \mu}{\sigma}\right) + (a - \mu)\left[2\Phi\left(\tfrac{a - \mu}{\sigma}\right) - 1\right].$$

Let $X \sim F$. Then:

$$\mathrm{E}|X - y| = \sum_{m=1}^{N} \omega_m A(\mu_m, \sigma_m; y),$$

and

$$\mathrm{E}|X - X'| = \sum_{m=1}^{N}\sum_{n=1}^{N} \omega_m \omega_n A(\mu_m - \mu_n, \sqrt{\sigma_m^2 + \sigma_n^2}).$$

Thus,

$$\mathrm{CRPS}(F, y) = \sum_{m=1}^{N} \omega_m A(\mu_m, \sigma_m; y) - \tfrac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N} \omega_m \omega_n A(\mu_m - \mu_n, \sqrt{\sigma_m^2 + \sigma_n^2}).$$

(c) Using the alternative expression of the CRPS, i.e.,

$$\mathrm{CRPS}(F, y) = \mathrm{E}_F(|X - y|) - \tfrac{1}{2}\mathrm{E}_F(|X - X'|),$$

we observe that the first term measures how close the forecast is to the true observation which tells us about its calibration, while the second term captures the spread (and hence the sharpness) of the forecast. For instance, if the forecast is very sharp (say, a point mass), then $|X - X'|$ will be very small. However, if the forecast is highly dispersive (say, with both tails being very heavy), that quantity can take large values.