Q: Show that the variance of a sum is $\mathrm{var}[X+Y]=\mathrm{var}[X]+\mathrm{var}[Y]+2\mathrm{cov}[X,Y]$.

A: From the definition of variance

$$
\begin{aligned}
\mathrm{var}[X+Y] =& \mathrm{E}[(X+Y-\mathrm{E}[X+Y])^2] \\
=& \mathrm{E}[X^2+Y^2+2XY-\mathrm{E}^2[X]-\mathrm{E}^2[Y]-2\mathrm{E}[X]\mathrm{E}[Y]] \\
=& \mathrm{E}[X^2-\mathrm{E}^2[X]]+\mathrm{E}[Y^2-\mathrm{E}^2[Y]]+2(\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]) \\
=& \mathrm{var}[X]+\mathrm{var}[Y]+2\mathrm{cov}[X,Y].
\end{aligned}
$$

Q: Assume $\theta \sim Beta(\alpha,\beta)$, derive the mean, mode and variance.

A: The probability density function (pdf) of θ is

$$
f(\theta;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},
$$

then the mean is

$$
\begin{aligned}
\mathrm{E}[\theta] &= \int_0^1 \theta f(\theta;\alpha,\beta)\mathrm{d}\theta \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)}\int_0^1 \frac{\theta^{\alpha}(1-\theta)^{\beta-1}}{B(\alpha+1,\beta)}\mathrm{d}\theta \\
&= \frac{\alpha}{\alpha+\beta},
\end{aligned}
$$

and the variance is

$$
\begin{aligned}
\mathrm{var}[\theta] =& \mathrm{E}[\theta^2]-\mathrm{E}^2[\theta] \\
=& \frac{(\alpha+1)\alpha}{\alpha+\beta+1}-\frac{\alpha^2}{(\alpha+\beta)^2} \\
=& \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}
$$

For the mode, because the derivative is

$$
f'(\theta;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-2}(1-\theta)^{\beta-2}[(\alpha-1)(1-\theta)-(\beta-1)\theta],
$$

when $f'=0$, it has $\theta=(\alpha-1)/(\alpha+\beta-2)$, which exists when $\alpha>1$ and $\beta>1$.

Q: show that a necessary and sufficient condition for Σ to be positive definite is that all the eigenvalues λ_i of Λ are positive.

A: Because $\Sigma=U^T\Lambda U$, if x is a real vector with the same dimension as Σ, then $x^T\Sigma x=(Ux)^T\Lambda(Ux)$.

Necessity: If Σ is a positive-definite matrix, then

$$
x^T\Sigma x=(Ux)^T\Lambda(Ux)>0 \text{ holds for any } x\neq 0.
$$

then $\lambda_i>0$, otherwise there exist x s.t. $(Ux)^T\Lambda(Ux)\leq 0$.

Sufficiency: If $\lambda_i > 0$, then

$$\begin{aligned}(Ux)^T\Lambda(Ux) =&(Ux)^T QAQ^{-1}(Ux)\\ =&\det A\cdot(Ux)^T(Ux) > 0,\end{aligned}$$

where QAQ^{-1} is the eigenvalue-decomposition of Λ. So Σ is a positive-definite matrix,

Q: Derive the maximum likelihood solutions for the mean and the variance of a univariate Gaussian.

A: The likelihood function is

$$L(\mu,\sigma^2;x_1,\cdots,x_n) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n(x-\mu)^2).$$

So the log likelihood is

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2.$$

MLE for μ:

$$\frac{\partial \ln L}{\partial\mu} = 0 \to \mu = \frac{1}{n}\sum_{i=1}^n x_i.$$

MLE for σ^2:

$$\frac{\partial \ln L}{\partial\sigma^2} = 0 \to \sigma^2 = \frac{1}{n}\sum_{i=1}^n(x_i-\hat{\mu})^2,$$

where $\hat{\mu}$ is the MLE for μ.

Q: Plot Gaussian likelihoods with unknown means, conjugate priors of the means, and their corresponding posterior distributions...

A: The conjugate prior $N(\mu \mid 0, 6)$,

The likelihood $p(D \mid \mu,\sigma^2)$ is as the answer in the above question. Specifically, here it is

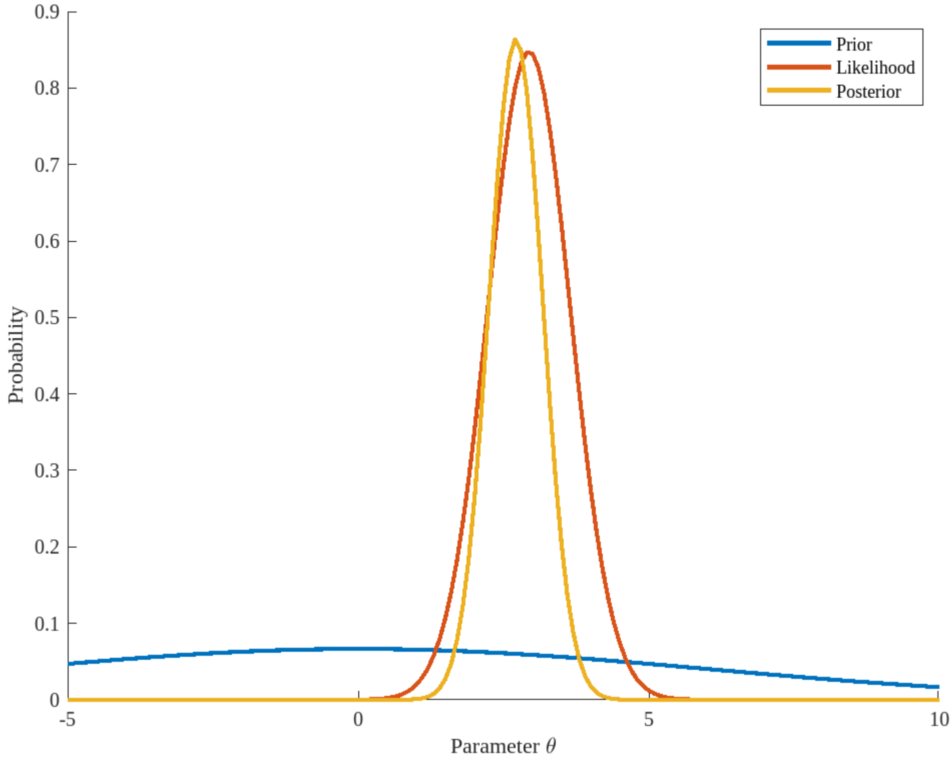
$$p(D\mid\mu,\sigma^2) = (2\pi\sigma^2)^{-10}\exp(-\frac{1}{2\sigma^2}\cdot 20s^2 + 20(\hat{x}-\mu)^2),$$

where s^2 is the empirical variance, $\sigma^2 = 10$, and the sample number is 20 in D.

The posterior is a Gaussian distribution with

$$\sigma_N^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2+N\sigma_0^2} = \frac{6}{13} \text{ and } \mu_N = \frac{N\sigma_0^2}{\sigma^2+N\sigma_0^2}\hat{x} + \frac{\sigma_0^2}{\sigma^2+N\sigma_0^2}\mu_0 = \frac{12}{13}\times 2.9390 = 2.7130.$$

The three plots are shown below:



The posterior predictive is given by $N(\mu_N, \sigma_N^2 + \sigma^2)$, so the probability of $x^* = 2.4$ is $N(2.4 \mid 2.7130, 6/13 + 10) = 0.0381$.

Q: Scalar QDA...

A: Denote the two classes "m" and "f" by 0 and 1. The conditional likelihood of the of each class is

$$\begin{aligned} p(x|y=0) =& N(x|\mu_0, \sigma_0) = |\sigma_0|^{-1/2} \exp(-1/2(x-\mu_0)^T \sigma_0^{-1}(x-\mu_0)) \\ p(x|y=1) =& N(x|\mu_1, \sigma_1) = |\sigma_1|^{-1/2} \exp(-1/2(x-\mu_1)^T \sigma_1^{-1}(x-\mu_1)), \end{aligned}$$

where $\mu_0 = 217/3$, $\sigma_0 \simeq 24.8889$, $\mu_1 = 65$, $\sigma_0 \simeq 12.6667$.

The prior of each class is $\pi_0 = \pi_1 = 1/2$.

For the predictive likelihood, the shared covariance is need, which is $\sigma = 32.2222$; then

$$p(y=0|x=72, \hat{\theta}) = \frac{p_0}{p_{01}},$$

where

$$\begin{aligned} p_0 =& |\sigma|^{-1/2} \exp(-\frac{1}{2}(x-\mu_0)^T \sigma^{-1}(x-\mu_0))\pi_0 \\ p_{01} =& p_0 + |\sigma|^{-1/2} \exp(-\frac{1}{2}(x-\mu_1)^T \sigma^{-1}(x-\mu_1))\pi_1. \end{aligned}$$

The result is $p(y=0|x=72, \hat{\theta}) = 0.6811$.