1. For \( n = 1, 2, 3, \ldots \) the joint probability mass function of the random variables \( x_1, x_2, \ldots, x_n \) is \( p_n(x_1, \ldots, x_n) \). These distributions are specified arbitrarily: the marginal distributions might be inconsistent for different \( n \).

(a) Suppose now \( x_1, x_2, x_3, \ldots \) is an infinite sequence of exchangeable binary random variables. Show that the marginal distributions must be consistent, i.e., show that
\[
p_n(x_1, \ldots, x_n) = p_{n+1}(x_1, \ldots, x_n, 0) + p_{n+1}(x_1, \ldots, x_n, 1).
\]

Remark: such families are “marginally consistent”. More generally we require
\[
p_n(x_1, \ldots, x_n) = \int p_{n+1}(x_1, \ldots, x_n, x_{n+1}) \, dx_{n+1}.
\]

(b) Show that \( \text{cov}(x_i, x_j) \geq 0 \) for all \( i, j \in \{1, 2, 3, \ldots\} \).

(c) Construct a prior for \( x_1, x_2, x_3, \ldots \) representing the following prior knowledge: (A) the variables are exchangeable; (B) \( \Pr(x_i = 1) = p \); and (C) \( \text{var}(\bar{x}) = v \) where \( \bar{x} = n^{-1} \sum_{i=1}^n x_i \) and \( 0 \leq p \leq 1 \) and \( v \) are prior parameters we wish to specify separately. Note any constraints on \( p \) and \( v \).

Hint: this is just Q1 of sheet 2 again.

2. Let \( X_1, X_2 \) be binary random variables. Table entries below give probabilities, \( p(x_1, x_2) = \Pr(X_1 = x_1, X_2 = x_2) \), for outcomes \( (X_1, X_2) = (x_1, x_2) \) indicated by row and column.

\[
\begin{array}{c|cc}
X_2 & X_1 = 0 & X_1 = 1 \\
\hline
X_2 = 0 & 0 & 1/2 \\
X_2 = 1 & 1/2 & 0 \\
\end{array}
\]

(a) Show that \( X_1 \) and \( X_2 \) are exchangeable.

(b) Show that there does not exist a distribution \( F \) such that
\[
p(x_1, x_2) = \int_0^1 \prod_{i=1,2} p^{x_i}(1-p)^{1-x_i} \, dF(p).
\]

Remark: i.e., de Finetti’s theorem need not hold if the sequence is finite. This example comes from a paper by Diaconis and Freedman (1980)

3. Let \( H \) be a distribution on \( \Omega \) and suppose \( G \sim DP(\alpha, H) \) with \( \alpha > 0 \) a real parameter.

(a) Let \( A \subset \Omega \). Calculate \( \text{var}(G(A)) \). Briefly interpret \( \alpha \) and \( H \) as model “parameters”.

(b) Suppose for \( i = 1, 2, 3, \ldots, \theta_i \sim G \) with \( G \sim DP(\alpha, H) \). Show that the predictive distribution of \( \theta_{n+1} \) given \( \theta_{1:n} = (\theta_1, \ldots, \theta_n) \), \( \alpha \) and \( H \) is
\[
(\theta_{n+1} \mid \theta_{1:n}, \alpha, H) \sim \frac{\alpha H + \sum_{i=1}^n \delta_{\theta_i}}{\alpha + n}.
\]
4. Consider the Dirichlet-Normal mixture with \( \theta = (\mu^*_i, \sigma^*_i) \) and for \( i = 1, \ldots, n \),
\[
y_i \sim N(\mu^*_i, \sigma^*_i)
\]
where for \( k = 1, \ldots, K \) and \( \theta^*_k = (\mu^*_k, \sigma^*_k) \)
\[
H(d\theta^*_k) = N(d\mu^*_k; \mu_0, \sigma_0)\Gamma(d\sigma^*_k; \alpha_0, \beta_0)
\]
in terms of the \( S, \theta^* \) notation of Q3c.

(a) Show that the prior expected number of clusters \( E(K) \) is a function of the number of observations \( n \), and that it diverges to infinity with \( n \).

(b) Write down the posterior \( \pi(S, \theta^*|y) \propto f(y|S, \theta^*)\pi(\theta^*)P(S) \) for \( S, \theta^*|y \) in terms of the model elements. Give the marginal likelihood \( p(y) \) in terms of \( f, \pi \) and \( P \).

(c) The \( \sigma \)-prior is not conjugate. Outline an MCMC algorithm targeting \( \pi(S, \theta^*|y) \).

(d) (optional) Modify the L16 MCMC code to target \( \pi(S, \theta^*|y) \) with \( \alpha_0 = 1.5, \beta_0 = 0.5 \).

5. Consider a prior for a normal mixture like the RJ-MCMC prior in lecture 11.
\[
M \sim \text{Geom}(\xi)
\]
\[
w \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_M) \quad \text{with} \quad \alpha_m = \alpha/M, \ m = 1, \ldots, M
\]
\[
z_i \sim \text{Multinom}(w), \ i = 1, \ldots, n
\]
\[
(\mu^*_m, \sigma^*_m) \sim N(\mu^*_m; \mu_0, \sigma_0)\Gamma(\sigma^*_m; \alpha_0, \beta_0) \ m = 1, \ldots, M
\]
Here \( z_i \sim \text{Multinom}(w), \ i = 1, \ldots, n \) means \( z_i = m, \ m \in \{1, \ldots, M\} \) with probability \( w_m \).

In this model \( z_i \) is the label of the cluster to which \( y_i \) belongs. The observation model is
\[
y_i \sim N(\mu^*_z, \sigma^*_z), \ i = 1, \ldots, n.
\]

(a) Suppose the list \( z_1, \ldots, z_n \) of cluster labels contains \( K \leq M \) unique values \( m_1, \ldots, m_K \).
For \( k = 1, \ldots, K \) let \( S_k = \{i : z_i = m_k, i = 1, \ldots, n\} \). Show that \( E(K) \leq 1/\xi \).

(b) Let \( S = (S_1, \ldots, S_K) \). Show that in this prior model
\[
P(S|M) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/M)^K} \frac{M!}{(M-K)!} \prod_{k=1}^K \frac{\Gamma(\alpha/M + n_k)}{\Gamma(\alpha + n)}
\]

(c) Write down the posterior \( \pi(S, \theta^*, M|y) \propto f(y|S, \theta^*)\pi(\theta^*)P(S|M)\pi_M(M) \) for \( S, \theta^*, M|y \) in terms of the model elements.

(d) Show (or at least outline why it is unsurprising) that the posterior converges to the DP posterior as \( \xi \to 0 \). Note that \( z\Gamma(z) \to 1 \) as \( z \to 0 \).