**Paso 1. Definición de Objetivo y Métrica**

1. Define claramente **qué quieres mejorar o probar**.
2. Define la **métrica principal** que guiará la decisión (ej. tasa de éxito, ingresos por usuario, tiempo de interacción).
3. Define métricas secundarias si quieres monitorear efectos colaterales.
4. Establece el **umbral de certeza** (ej.: 95%) y lift mínimo relevante (ej.: +1 pp o +$X por usuario).
5. Define valor económico de la métrica si deseas usar **expected loss** o **funciones de utilidad**.

**Opciones/metodologías que se pueden usar después:**

* Probabilidad de superioridad (P(B>A)).
* Intervalos de credibilidad (Bayesiano equivalente a intervalos de confianza).
* Expected loss / funciones de utilidad (para decisiones económicas o riesgosas).

**Paso 2. Planteamiento del problema y distribución a priori**

1. Recolecta históricos: total eventos n\_hist y éxitos k\_hist.
2. Define la **distribución a priori**:
   * Para métricas binarias: Beta(α, β).
   * Para métricas continuas: Gaussianas o distribuciones conjugadas.
3. Escoge priors según conocimiento:
   * Basados en datos históricos.
   * Priors débiles/no informativos si no hay datos (Beta(1,1) o Normal(0,1)).
   * Priors jerárquicos si hay segmentación o múltiples variantes.
4. Documenta la elección de priors y justificación.

**Opciones/metodologías que se pueden usar después:**

* Priors informativos vs no informativos.
* Priors jerárquicos (para segmentos o múltiples variantes).
* Priors robustos con “discounting” de datos históricos antiguos.

**Paso 3. Diseño experimental**

1. Define variantes (ej.: A = control, B = tratamiento).
2. Define proporción de asignación (50/50 por defecto).
3. Garantiza que **usuarios no se mezclen** entre grupos.
4. La **muestra**:
   * No es fija, se acumula con el tiempo.
   * Se puede estimar **muestra mínima orientativa** con:
     + Fórmulas clásicas (statsmodels, efecto esperado).
     + Simulación forward (Monte Carlo) para estimar número de eventos necesarios para alcanzar certeza.
5. Define reglas de monitoreo diario y métricas de control de calidad.

**Opciones/metodologías que se pueden usar:**

* Tamaño de muestra clásico (frecuentista) como referencia.
* Simulación forward (Monte Carlo) para estimar duración y volumen necesario.
* Multi-arm bandit / Thompson Sampling si quieres ir ajustando la asignación en tiempo real.

**Paso 4. Ejecución y recolección de datos**

1. Registrar eventos con campos mínimos: id, grupo, timestamp, outcome, intento, metadata.
2. Monitoreo diario:
   * Conteo de eventos por grupo.
   * Chequeo de asignación correcta y datos faltantes.
3. No cambiar usuarios entre grupos.
4. La recolección es continua hasta que se cumpla la regla de parada.

**Opciones/metodologías que se pueden usar:**

* Actualización de posterior diaria.
* Monitoreo en tiempo real para bandit/Thompson Sampling.
* Segmentación dinámica: calcular posterior por segmentos y agregarlos jerárquicamente.

**Paso 5. Actualización de la distribución posterior**

1. Para métricas binarias: α\_post = α\_prior + k\_observed, β\_post = β\_prior + n\_observed - k\_observed.
2. Para métricas continuas: actualizar media y desviación usando conjugación normal o PyMC.
3. Guardar posterior para cada snapshot.
4. Repetir cada vez que haya nuevos datos.

**Opciones/metodologías que se pueden usar:**

* Fórmulas exactas (conjugadas).
* Monte Carlo: simular muestras del posterior para cálculo flexible de probabilidades.
* Modelos jerárquicos (PyMC) si hay segmentos o múltiples variantes.
* Thompson Sampling: usar posterior para decidir asignación dinámica.

**Paso 6. Análisis de resultados**

1. Métodos básicos:
   * Probabilidad de superioridad: P(B>A).
   * Intervalos de credibilidad para la métrica y lift.
2. Métodos avanzados:
   * Monte Carlo: generar muchas muestras de cada posterior.
   * Expected loss: calcular pérdida esperada si eliges una variante equivocada.
   * Funciones de utilidad: maximizar valor esperado en lugar de minimizar pérdida.
3. Visualización:
   * Densidad posterior por variante.
   * Histograma o densidad del lift (B−A).
   * Tabla con P(B>A), lift esperado, intervalos y expected loss.

**Opciones/metodologías que se pueden usar:**

* Simulación forward para estimar tiempo/muestras requeridas.
* Comparación directa de distribución posterior con múltiples variantes.
* Análisis por segmento o covariable.

**Paso 7. Decisión**

1. Define reglas de decisión:
   * Si P(B>A) ≥ 95% y lift relevante → elegir B.
   * Si expected loss de B es mínimo y aceptable → elegir B.
   * Si reglas no se cumplen y se llega a T\_max/N\_max → cerrar como inconcluso.
2. Documenta la decisión: parámetros posteriores, lift, expected loss, probabilidad de superioridad.
3. Implementa la variante ganadora y planifica **rollout**.
4. Actualiza posterior final como prior para futuros experimentos.

**Opciones/metodologías que se pueden usar:**

* Regla basada solo en P(B>A) (básico).
* Regla basada en intervalos de credibilidad.
* Expected loss monetario o en puntos de métrica.
* Funciones de utilidad para decisiones más estratégicas.
* Thompson Sampling si decides seguir optimizando tráfico en tiempo real.